

08/18/00  
JC907 U.S. PTO

8-21-00

A

# FISH & RICHARDSON P.C.

225 Franklin Street  
Boston, Massachusetts  
02110-2804

Telephone  
617 542-5070

Facsimile  
617 542-8906

Web Site  
[www.fr.com](http://www.fr.com)

Frederick P. Fish  
1855-1930

W.K. Richardson  
1859-1951

August 18, 2000

Attorney Docket No.: 11910-002001

jc873 U.S. PTO  
09/641589  
08/18/00

**Box Patent Application**  
Commissioner for Patents  
Washington, DC 20231

Presented for filing is a new original patent application of:

BOSTON  
DALLAS  
DELAWARE  
NEW YORK  
SAN DIEGO  
SILICON VALLEY  
TWIN CITIES  
WASHINGTON, DC  
JC907 U.S. PTO  
09/641589  
08/18/00

Applicant: PHILIP A. COOPER, LISETTE COOPER, STEWART MYERS,  
G. DAVID FORNEY, JR., AND LEONARD L. SCOTT, JR.

Title: GENERATING AND PROVIDING INFORMATION ABOUT  
EXPECTED FUTURE PRICES OF ASSETS

Enclosed are the following papers, including those required to receive a filing date  
under 37 CFR §1.53(b):

	<u>Pages</u>
Specification	29
Claims	8
Abstract	1
Declaration	[To be Filed at a Later Date]
Drawing(s)	8

Enclosures:

- Postcard.
- Appendix A (18 pages).

## CERTIFICATE OF MAILING BY EXPRESS MAIL

Express Mail Label No. EL485673071US

I hereby certify under 37 CFR §1.10 that this correspondence is being  
deposited with the United States Postal Service as Express Mail Post  
Office to Addressee with sufficient postage on the date indicated below  
and is addressed to the Commissioner for Patents, Washington,  
D.C. 20231.

August 18, 2000  
Date of Deposit

Signature

Samantha Bell  
Typed or Printed Name of Person Signing Certificate

FISH & RICHARDSON P.C.

Commissioner for Patents  
August 18, 2000  
Page 2

Basic filing fee	\$345
Total claims in excess of 20 times \$9	\$63
Independent claims in excess of 3 times \$39	\$546
Fee for multiple dependent claims	\$0
Total filing fee:	\$954

Under 37 CFR §1.53(f), no filing fee is being paid at this time.

If this application is found to be incomplete, or if a telephone conference would otherwise be helpful, please call the undersigned at (617) 542-5070.

Kindly acknowledge receipt of this application by returning the enclosed postcard.

Please send all correspondence to:

DAVID L. FEIGENBAUM  
Fish & Richardson P.C.  
225 Franklin Street  
Boston, MA 02110-2804

Respectfully submitted,



David L. Feigenbaum  
Reg. No. 30,378  
Enclosures  
DLF/lja  
20079260.doc

DRAFT - DO NOT CITE

## **GENERATING AND PROVIDING INFORMATION ABOUT EXPECTED FUTURE PRICES OF ASSETS**

## BACKGROUND

This invention relates to generating and providing information about expected future prices of assets.

Among the kinds of information available at web sites on the Internet are current and historical prices and volumes of stock transactions, prices of put or call options at specific strike prices and expiration dates for various stocks, and theoretical prices of put and call options that are derived using formulas such as the Black-Scholes formula. Some web sites give predictions by individual experts of the future prices or price ranges of specific stocks.

A call option gives the holder a right to buy an underlying marketable asset by an expiration date for a specified strike price. A put option gives an analogous right to sell an asset. Options are called derivative securities because they derive their values from the prices of the underlying assets. Examples of underlying assets are corporate stock, commodity stock, and currency. The price of an option is sometimes called the premium.

People who buy and sell options are naturally interested in what appropriate prices might be for the options. One well-known formula for determining the prices for call and put options under idealized conditions is called the Black-Scholes formula. Black-Scholes provides an estimate of call or put prices for options having a defined expiration date, given a current price of the underlying asset, an interest rate, and the volatility rate (sometimes simply called volatility) of the asset. Black-Scholes assumes

constant interest rates and volatility, no arbitrage, and trading that is continuous over a specified price range.

## **SUMMARY**

In general, in one aspect, the invention features a method in which data is

5 received that represents current prices of options on a given asset. An estimate is derived from the data of a corresponding implied probability distribution of the price of the asset at a future time. Information about the probability distribution is made available within a time frame that is useful to investors, for example, promptly after the current option price  
10 information becomes available.

Implementations of the invention may include one or more of the following features. The data may represent a finite number of prices of options at spaced-apart strike prices of the asset. A set of first differences may be calculated of the finite number of prices to form an estimate of the  
15 cumulative probability distribution of the price of the asset at a future time. A set of second differences may be calculated of the finite number of strike prices from the set of first differences to form the estimate of the probability distribution function of the price of the asset at a future time.

In general, in another aspect, the invention features a method in which a

20 real time data feed is provided that contains information based on the probability distribution.

In general, in another aspect, the invention features a method that includes providing a graphical user interface for viewing pages containing financial information related to an asset; and when a user indicates an asset of  
25 interest, displaying probability information related to the price of the asset at a future time.

In general, in another aspect, the invention features a method that includes receiving data representing current prices of options on a given asset, the options being associated with spaced-apart strike prices of the asset at a future time. The data includes shifted current prices of options resulting  
5 from a shifted underlying price of the asset, the amount by which the asset price has shifted being different from the amount by which the strike prices are spaced apart. An estimate is derived from a quantized implied probability distribution of the price of the asset at a future time, the elements of the quantized probability distribution being more finely  
10 spaced than for a probability distribution derived without the shifted current price data.

In general, in another aspect, the invention includes deriving from said data an estimate of an implied probability distribution of the price of the asset at a future time, the mathematical derivation including a smoothing  
15 operation.

Implementations of the invention may include one or more of the following features. The smoothing operation may be performed in a volatility domain.

In general, in another aspect, the invention includes deriving a volatility  
20 for each of the future dates in accordance with a predetermined option pricing formula that links option prices with strike prices of the asset; and generating a smoothed and extrapolated volatility function.

Implementations of the invention may include one or more of the following features. The volatility function may be extrapolated to a wider  
25 range of dates than the future dates and to other strike prices. The smoothed volatility function may be applicable to conditions in which the data is reliable under a predetermined measure of reliability. The implied

volatility function formula may have a quadratic form with two variables representing a strike price and an expiration date. The coefficients of the implied volatility function formula may be determined by applying regression analysis to approximately fit the implied volatility function formula to each of the implied volatilities.

In general, in another aspect, the invention features a method that includes receiving data representing current prices of options on assets belonging to a portfolio, deriving from the data an estimate of an implied multivariate distribution of the price of a quantity at a future time that depends on the assets belonging to the portfolio, and making information about the probability distribution available within a time frame that is useful to investors.

In general, in another aspect, the invention features a method that includes receiving data representing values of a set of factors that influence a composite value, deriving from the data an estimate of an implied multivariate distribution of the price of a quantity at a future time that depends on assets belonging to a portfolio, and making information about the probability distribution available within a time frame that is useful to investors.

Implementations of the invention may include one or more of the following features. The mathematical derivation may include generating a multivariate probability distribution function based on a correlation among the factors.

In general, in another aspect, the invention features a graphical user interface that includes a user interface element adapted to enable a user to indicate a future time, a user interface element adapted to show a current

price of an asset, and a user interface element adapted to show the probability distribution of the price of the asset at the future time.

In general, in one aspect, the invention features, a method that includes continually generating current data that contains probability distributions 5 of prices of assets at future times, continually feeding the current data to a recipient electronically, and the recipient using the fed data for services provided to users.

In general, in another aspect, the invention features a method that includes 10 receiving data representing current prices of options on assets belonging to a portfolio, receiving data representing current prices of market transactions associated with a second portfolio of assets, and providing information electronically on the probability that the second portfolio of assets will reach a first value given the condition that the first portfolio of assets reaches a specified price at a future time.

15 In general, in another aspect, the invention features a method that includes receiving data representative of actual market transactions associated with a first portfolio of assets; receiving data representative of actual market transactions associated with a second portfolio of assets; and providing information on the expectation value of the price of first portfolio of assets 20 given the condition that the second portfolio of assets reach a first specified price at a specified future time through a network.

In general, in another aspect, the invention features a method that includes 25 evaluating an event defined by a first multivariate expression that represents a combination of macroeconomic variables at a time T, and estimating (e.g., using Monte Carlo techniques) the probability that a second multivariate expression that represents a combination of values of assets of a portfolio will have a value greater than a constant B at time T if

the value of the first multivariate expression is greater than a constant A. The market variables represented by the first multivariate expression can include macroeconomic factors (such as interest rates), market preferences regarding the style of company fundamentals (large/small companies, 5 rapid/steady growth, etc.), or market preferences for industry sectors.

In general, in another aspect, the invention features a method that includes defining a regression expression that relates the value of one variable representing a combination of macroeconomic variables at time T to a second variable at time T that represents a combination of assets of a 10 portfolio, and estimating the probability that the second variable will have a value greater than a constant B at time T if the value of the first variable is greater than a constant A at time T, based on the ratio of the probability of x being greater than A under the regression expression and the probability of x being greater than A.

15 In general, in another aspect, the invention features a method that includes defining a current value of an option as a quadratic expression that depends on the difference between the current price of the option and the current price of the underlying security, and using Monte Carlo techniques to estimate a probability distribution of the value at a future time T of a 20 portfolio that includes the option.

The invention takes advantage of the realization that option prices for a given underlying asset are indicative of the market's prediction of the of the risk-neutral price of the underlying asset in the future (e.g., at the 25 expiration of the option). Option price data may be used to derive the market's prediction in the form of an implied probability distribution of future risk-neutral prices. Additional explanation of the significance of the phrase risk-neutral is contained in the Appendix.

The implied probability distribution and other information related to it may be made easily available to people for whom the information may be useful, such as those considering an investment in the underlying asset, or a brokerage firm advising such an investor.

5 Among the advantages of the invention are one or more of the following: Investors and prospective investors in an underlying asset, such as a publicly-traded stock, are given access to a key additional piece of current information, namely calculated data representing the market's view of the future price of the stock. Brokerage firms, investment advisors, and other  
10 companies involved in the securities markets are able to provide the information or related services to their clients and customers.

Other features and advantages will become apparent from the following description and from the claims.

## **DESCRIPTION**

15 Details of implementations of the invention are set forth in the figures and the related description below.

Figures 1, 2, and 3 are graphs.

Figure 4 is a block diagram.

Figures 5, 6, and 7 are web pages.

20 Figures 8 and 9 illustrate user interfaces.

Figure 10 shows data structures.

In general, the price of a call or put option is determined by buyers and sellers in the option market and carries information about the market's

prediction of the expected price of the underlying asset at the expiration date. (The information does not include the premium that investors require for bearing risk, which must be estimated separately. The average long-term value of the risk premium is about 6% per year for all stocks and may 5 be adjusted for an individual stock's historical responsiveness to broader market movements.)

The information carried in the prices of options having various strike prices and expirations is used to derive probability distributions of the asset's price at future times and to display corresponding information to 10 investors, for example, on the World Wide Web.

#### *Basic method*

We first define some relevant terms. We define  $x$  as the strike price,  $c(x)$  as the theoretical call price function (the price of the call as a function of strike price),  $p(x)$  as the theoretical put price function,  $F(x)$  as the 15 cumulative distribution function (cdf) of the price of the underlying asset at expiration; and  $f(x)$  as the probability density function (pdf) of the asset price at expiration. By definition,  $f(x) = F'(x)$  (i.e., the probability density function is the derivative of the cumulative distribution function).

The relationship between  $c(x)$ ,  $p(x)$ ,  $f(x)$ , and  $F(x)$  can be succinctly stated

20 as:

$$F(x) = c'(x) + 1 = p'(x); \quad (1a)$$

$$f(x) = c''(x) = p''(x). \quad (1b)$$

In words, the pdf is the second derivative of either the call price function or of the put price function. A simple proof these relationships is given in 25 the Appendix. The Appendix also contains other detailed information relating to the features of the invention.

This so-called "second-derivative method" for computing implied probability distributions from option price data is known in the academic literature, but apparently not very well known. For example, the standard textbook "Options, Futures, and Other Derivatives," by John C. Hull

5 (Fourth Edition, 1999; Prentice-Hall) mentions implied probabilities, but not the second-derivative method. Perhaps the best reference that we have been able to find is J. C. Jackwerth and M. Rubinstein, "Recovering probability distributions from option prices," *J. Finance*, vol. 51, pp. 1611-1631 (1996), which has only six prior references. This paper cites D. T.

10 Breeden and R. H. Litzenberger, "Prices of state-contingent claims implicit in option prices," *J. Business*, vol. 51, pp. 631-650 (1978) as the originator of a second-derivative method, although the latter paper nowhere mentions probabilities.

*Approximating  $f(x)$  from finite bid and ask option prices*

15 Equations (1a) and (1b) are obtained by assuming that the variable  $x$  is continuous and ranges from 0 to infinity. In practice, options are usually traded within certain price ranges and only for certain price intervals (e.g., ranging from \$110 to \$180 at \$5 intervals). Thus, the call and/or put option prices are known only for a finite subset of strike prices. Under

20 such circumstances, estimates of Equations (1a) and (1b) can be computed by taking differences instead of derivatives as follows.

We assume that the option prices  $c(x)$  and  $p(x)$  are quoted for a finite subset of equally-spaced strike prices  $x = n \Delta$ , where  $n$  is an integer, and  $\Delta$  is the spacing between quoted prices. Define  $c_n = c(n \Delta)$ ,  $p_n = p(n \Delta)$ . Then

25 the first derivatives  $c'(x)$  and  $p'(x)$  at  $x = (n + \frac{1}{2})\Delta$  may be estimated by the first differences:

$$\hat{c}'_{n+1/2} = \frac{c_{n+1} - c_n}{\Delta}; \quad (7a)$$

$$\hat{p}'_{n+1/2} = \frac{p_{n+1} - p_n}{\Delta}. \quad (7b)$$

The corresponding estimates of the cumulative distribution function:

$$\hat{F}_{n+1/2} = F((n + \frac{1}{2})\Delta) \text{ are}$$

$$5 \quad \hat{F}_{n+1/2} = 1 + \hat{c}'_{n+1/2} \quad (8a)$$

$$\hat{F}_{n+1/2} = \hat{p}'_{n+1/2} \quad (8b)$$

The second derivatives  $c''(x)$  and  $p''(x)$  at  $x = n \Delta$  may likewise be estimated by the second differences, i.e., differences of the estimates of the first derivatives:

$$10 \quad \hat{c}''_n = \frac{c_{n+1/2} - c_{n-1/2}}{\Delta} = \frac{c_{n+1} - 2c_n + c_{n-1}}{\Delta^2}; \quad (9a)$$

$$\hat{p}''_n = \frac{p_{n+1} - 2p_n + p_{n-1}}{\Delta^2}. \quad (9b)$$

Either of these estimates of the second derivatives may be used as an estimate of the probability density values at  $x = n \Delta$ , i.e.,  $f(n\Delta)$ :

$$\hat{f}_n = \hat{p}''_n \text{ or } \hat{f}_n = \hat{c}''_n \quad (10)$$

15 Moreover, the market prices of call and put options are usually given in terms of a bid-ask spread, and thus either the bid price or the ask price (or some intermediate value) may be used as the call or put option price. By using the bid and ask prices for both the call option and the put option, four estimates of  $F(x)$  and  $f(x)$  may be obtained. These estimates may be

combined according to their reliability in any desired way. For example, one might use the estimate derived from the put bid price curve for values of  $x$  less than the current price  $s$  of the underlying asset, and the estimate derived from the call bid price curve for values of  $x$  greater than  $s$ .

5 Examples of  $c_n$ ,  $p_n$ ,  $\hat{F}_{n+1/2}$ , and  $\hat{f}_n$  are shown in figures 1, 2, and 3 using the data of TABLE 1 (see below).

*Tabular data*

TABLE 1 below shows sample bid prices of call and put options for strike prices of an asset ranging from \$110 to \$180 at \$5 intervals and the

10 cumulative distribution values  $\hat{F}_{n+1/2}$  and probability density values  $\hat{f}_n$  computed according to Equations (7)-(10) above.

In the table, the values for  $\hat{F}_{n+1/2}$  correspond to strike prices that are midway between the two strike prices used to compute  $\hat{F}_{n+1/2}$ . Thus, the cumulative distribution value shown to the right of the strike price \$110 actually corresponds to the strike price \$112.5, and the value to the right of the strike price \$115 actually corresponds to strike price \$117.5, and so forth.

Strike price	Call price	$\hat{F}_{n+1/2}$ from call price	$\hat{f}_n$ from call price	Put price	$\hat{F}_{n+1/2}$ from put price	$\hat{f}_n$ from put price
110	42 7/8	0	0	1/	0	0
115	37 7/8	0.025	0.025	1/	0.0125	0.0125
120	33	0.025	0	3/1	0.0125	0
125	28 1/8	0.05	0.025	1/	0.0375	0.025
130	23 3/8	0.125	0.075	7/1	0.0875	0.05
135	19	0.175	0.05	7/	0.15	0.0625
140	14 7/8	0.225	0.05	1 5/	0.2375	0.0875
145	11	0.35	0.125	2 13/1	0.3875	0.15
150	7 3/4	0.525	0.175	4 3/	0.5	0.1125
155	5 3/8	0.6	0.075	7 1/	0.6	0.1
160	3 3/8	0.7375	0.1375	10 1/	0.725	0.125
165	2 1/16	0.825	0.0875	13 7/	0.825	0.1
170	1 3/16	0.8875	0.0625	1	0.85	0.025
175	5/8	0.925	0.0375	22 1/	0.925	0.075
180	1/4			26 7/		

Dynamic estimates for  $F(x)$  and  $f(x)$ .

In Equations (7)-(10), the call and put option prices were assumed to be static in the calculation of the cumulative distribution function  $F(x)$  and probability density function  $f(x)$  for a finite subset of strike prices  $x = n \Delta$ .  
5 In the real world, the price  $s$  of the underlying asset changes with time, and there will be a corresponding change in option prices. As a first order approximation, if the price  $s$  increases by a small amount  $\delta$ , then the option price curves will effectively shift to the right by the amount  $\delta$ .  
10 (Here,  $\delta$  may be either positive or negative. For a more precise discussion of the shift, see the Appendix.) As a result, the price  $c(x)$  or  $p(x)$  now quoted at strike price  $x$  may be used as an estimate for the option price on

the previous price curve at strike price  $x' = x - \delta$ . As a result, the prices on the previous curve at a new discrete subset of strike prices  $x = n\Delta - \delta$  become effectively visible. Given enough movements of the underlying price, therefore, we can effectively compute estimates of  $c(x)$ ,  $p(x)$ ,  $F(x)$  and/or  $f(x)$  for a subset of strike prices  $x$  that is much more closely spaced than the subset available at any one time.

*Extrapolating and smoothing probability distributions.*

In a typical options market, the option prices are available only for certain expiration dates. In addition, the option prices are more reliable for options that are actively traded, which are typically nearer-term options at strike prices near the underlying price. It is therefore desirable to extrapolate and interpolate probability distributions to times other than actual expiration dates and to wider ranges of strike prices.

Any standard extrapolation and smoothing techniques may be used directly on the cumulative distribution values  $\hat{F}_{n+1/2}$  or probability density values  $\hat{f}_n$  to give a smoothed and extrapolated estimate of  $F(x)$  or  $f(x)$ . Similarly, given such estimated curves for a discrete subset of future times  $T$ , standard interpolation and extrapolation techniques may be used to estimate such curves for other specified values of  $T$ , or for a continuous range of  $T > 0$ .

A less direct but useful approach is to perform extrapolation and smoothing on an implied volatility function, which is then used to calculate the other functions, such as  $c(x)$ ,  $p(x)$ ,  $F(x)$ , and  $f(x)$ . The volatility rate of an asset (often simply called its volatility) is a measure of uncertainty about the returns provided by the asset. The volatility rates of a stock may typically be in the range of 0.3 to 0.5 per year.

An advantage of performing extrapolation and smoothing on implied volatility curves is that different types of volatility curves (so-called “volatility smiles”) are known and can be used as a guide to the extrapolation and smoothing process to prevent “overfitting” of certain 5 unreliable data points.

The standard method of computing implied volatilities is to invert the Black-Scholes pricing formula (see Appendix) for the actual call price  $c(x)$  or put price  $p(x)$  of an underlying asset at a given strike price  $x$ , given the underlying price  $s$  (current price of asset), risk-free rate of interest  $r$ , and 10 and  $T$  (expiration date). When this is done for a range of values of  $x$ , an estimate of an implied volatility curve  $\sigma(x)$  is obtained. This curve may be smoothed and extrapolated by any standard method to give a smoothed curve  $\tilde{\sigma}(x)$ . Then corresponding smoothed put and call price curves may be computed using the Black-Scholes pricing formula and differentiated 15 once or twice to give a smoothed cdf or pdf. Finally, given such estimated curves for a discrete subset of future times  $T$ , standard interpolation and extrapolation techniques may be used to estimate such curves for other specified values of  $T$ , or for a continuous range of  $T > 0$ .

Another new way to compute implied volatilities is first to compute a 20 finite subset of cdf values  $\hat{F}_{n+1/2}$  and then to invert the Black-Scholes cdf formula (see Appendix) at these values. When this is done for a range of values of  $x$ , an estimate of a generally different implied volatility curve  $\sigma_l(x)$  is obtained, called the cdf-implied volatility curve. Again, this curve may be smoothed and extrapolated by any standard method to give a 25 smoothed curve  $\tilde{\sigma}_l(x)$ . Then a corresponding smoothed cdf may be computed from the Black-Scholes cdf formula, and differentiated once to give a smoothed pdf. Finally, again, given such estimated curves for a discrete subset of future times  $T$ , standard interpolation and extrapolation

techniques may be used to estimate such curves for other specified values of  $T$ , or for a continuous range of  $T > 0$ .

Some advantages of using the cdf-implied volatility curve rather than the conventional implied volatility curve are that the computations are

5 simpler, at least from an estimate of  $F(x)$ , and that it fits better with the multivariate techniques to be discussed below.

A particular method for finding a smoothed and extrapolated implied volatility curve  $\tilde{\sigma}_1(x, T)$  as a function of both strike price  $x$  and time  $T$  to expiration is as follows. The volatility curve is assumed to be

10 approximated by a quadratic formula

$$\tilde{\sigma}_1(x, T) = a_0 + a_1 x + a_2 x^2 + a_3 T + a_4 T^2 + a_5 x T, \quad (14)$$

The coefficients  $\{a_i\}$  are determined by regression to fit the available data regarding  $\sigma_1(x, T)$  as closely as possible. Given the smoothed curve  $\tilde{\sigma}_1(x, T)$ , corresponding smoothed cdfs for different  $x$ 's and  $T$ 's) may be

15 computed from the Black-Scholes cdf formula for each time  $T$ , and differentiated once to give a smoothed pdf. An alternative procedure, with numerical advantages, is to use a quadratic fit like the above for a function  $\tilde{\sigma}(x, T)$ , and then invert the Black-Scholes cdf to find  $\tilde{\sigma}_1(x, T)$ . See the Appendix for the academic history of such approximations of  $\tilde{\sigma}(x, T)$ .

20 Another useful variation is to fit  $\tilde{\sigma}(x, T)$  with a quadratic function of  $x$  at times  $T$  which are specific expiration dates, then linearly interpolate at other times  $T$ .

### *Treatment of multiple assets*

The techniques described so far give probability distributions for the

25 future values of a single asset based on option price data for that asset. However, in many cases an investor may be concerned with multiple

assets, for example all of the stocks in his or her portfolio, or in a mutual fund, or in a certain index. Moreover, the investor may be concerned with the relations between one group of assets and another.

A general method for dealing with such questions is to generate  
5 multivariate probability distributions for all assets of interest. A  
multivariate cdf may be written as  $F(x_1, x_2, \dots, x_n)$ , where the variables  $(x_1, x_2, \dots, x_n)$  are the values of the  $n$  assets of interest.

We will assume that we know from the techniques described above or  
otherwise the marginal cdfs  $F_i(x_i)$  for each of the individual variables. As  
10 a first step, we may define for each  $x_i$  a function  $y_i(x_i)$ , called a “warping  
function,” such that  $y_i(x_i)$  is a standard normal (Gaussian) variable with  
mean 0 and variance 1. This is simply done by defining  $y_i(x_i)$  such that  
 $F_i(x_i) = N(y_i(x_i))$  for all values of  $x$ , where  $N(x)$  denotes the cdf of a  
standard normal variable. The function  $y_i(x_i)$  may be simply described in  
15 terms of  $\sigma_i(x_i)$ . See the Appendix. Under mild technical conditions such  
as having a marginal cdf that varies monotonically, such a warping  
function  $y_i(x_i)$  has a well-defined inverse warping function  $x_i(y_i)$ .

Second, we assume that we can find the historical pairwise correlations  
between the warped standard normal variables  $y_i(x_i)$ . These correlations  
20 may be computed by standard techniques from any available set of  
historical asset price data. We denote by  $C$  the  $n \times n$  correlation matrix  
whose entries are these historically-based correlations. Because each of  
the variables  $y_i(x_i)$  is standard normal, the diagonal terms of  $C$  are all equal  
to 1.

25 Now let  $F_C(x_1, \dots, x_n)$  denote the cdf of a multivariate Gaussian random  $n$ -tuple with zero mean and covariance matrix  $C$ . Define

$$F(x_1, x_2, \dots, x_n) = F_C(y_1(x_1), y_2(x_2), \dots, y_n(x_n))$$

Then  $F(x_1, x_2, \dots, x_n)$  is a multivariate cdf that (a) has the correct (given) marginal cdfs  $F_i(x_i)$ ; and (b) has the correct (historical) correlations between the warped standard normal variables  $y_i(x_i)$ . We use this cdf to

5 answer questions involving the variables  $(x_1, x_2, \dots, x_n)$ .

For example, the investor might have a portfolio consisting of a given quantity of each of these assets. The value of such a portfolio is the sum

$$x = h_1 x_1 + h_2 x_2 + \dots + h_n x_n, \quad (15)$$

where  $h_i$  represents the quantity of the  $i$ th asset in the portfolio. The

10 investor might be interested in an estimate of the probability distribution of the value  $x$  of the whole portfolio.

Such an estimate may be obtained by Monte Carlo simulation. For such a simulation, a large number  $N$  of samples from the multivariate Gaussian cdf  $F_C(y_1, \dots, y_n)$  may be generated. Each sample  $(y_1, \dots, y_n)$  may be converted to a sample  $(x_1, x_2, \dots, x_n)$  by using the inverse warping functions  $x_i(y_i)$ . The value  $x$  of the total portfolio may then be computed for each sample. From these  $N$  values of  $x$ , the probability distribution of  $x$  (e.g., its cdf  $F(x)$ ) may be estimated.

15 In practice, it is useful to save the  $N$  multivariate samples in a large database. Then the cdf of any quantity whose value is a function of the variables  $(x_1, x_2, \dots, x_n)$  may be estimated from this database. For example, if the investor would like to know the cdf of some alternative portfolio with different quantities of each asset, this can be quickly determined from the stored database.

An investor may also determine the effect of one portfolio (or event(s) or variables such as interest rates, P/E ratios, public interest in a certain sector of the market) on another portfolio as follows. Assume that the first portfolio is represented by  $x$ , where

$$5 \quad x = h_1 x_1 + h_2 x_2 + \dots + h_n x_n, \quad (30)$$

where each  $x_i$  may be viewed as the price of a portfolio component, and the second portfolio is represented by  $y$ , where

$$y = g_1 x_1 + g_2 x_2 + \dots + g_n x_n. \quad (31)$$

where each  $y_i$  may be viewed as the price of a portfolio component or

more broadly as any macro-economic variable (macroeconomic, fundamental, or sector related).

Consider the “what-if” question: letting  $A$  and  $B$  be given positive constants, if  $x \geq A$  at time  $T$ , what is the probability that  $y \geq B$  at time  $T$ ?

This question can be answered by creating a Monte Carlo database as

above for the multivariate cdf  $F(x_1, x_2, \dots, x_n)$  corresponding to time  $T$ , identifying those samples for which  $x \geq A$ , and then using only these samples to estimate the probability that  $y \geq B$ . More generally, any conditional cdf of the form  $F(x | E)$  can be estimated similarly, where  $x$  is any function of the variables  $(x_1, x_2, \dots, x_n)$  and  $E$  is any event defined in terms of the variables  $(x_1, x_2, \dots, x_n)$ .

Similarly, suppose an investor would like to know whether it is reasonable to believe that a certain stock or portfolio  $x$  will have a value greater than a given constant  $A$  at time  $T$ . This kind of question can be addressed by estimating the conditional cdf of some other related and perhaps better-

25 understood variable (or combination of variables)  $y$  at time  $T$ , given that  $x$

$\geq A$ . If the resulting distribution for  $y$  does not look reasonable, then the investor may conclude that it is unreasonable to expect that  $x \geq A$ .

*Applications that use the probability distribution information*

A wide variety of techniques may be used to accumulate and process the information needed for the calculations described above and to provide the information to users directly or indirectly through third parties. Some of these techniques are described below.

As shown in figure 4, the probability distribution information can be provided to users from a host server 102 connected to a communication network 104, for example, a public network such as the Internet or a private network such as a corporate intranet or local area network (LAN). For purpose of illustration, the following discussion assumes that network 104 is the Internet.

The host server 102 includes a software suite 116, a financial database 120, and a communications module 122. The communications module 122 transmits and receives data generated by the host server 102 according to the communication protocols of the network 104.

Also connected to the network are one or more of each of the following (only one is shown in each case): an individual or institutional user 108, an advertisement provider 110, a financial institution 112, a third party web server 114, a media operator 122, and a financial information provider 106.

The operator of the host server could be, for example, a financial information source, a private company, a vendor of investment services, or a consortium of companies that provides a centralized database of information.

The host server 102 runs typical operating system and web server programs that are part of the software suite 116. The web server programs allow the host server 102 to operate as a web server and generate web pages or elements of web pages, e.g., in HTML or XML code, that allow 5 each user 108 to receive and interact with probability distribution information generated by the host server.

Software suite 116 also includes analytical software 118 that is configured to analyze data stored in the financial database 120 to generate, for example, the implied probability distribution of future prices of assets and 10 portfolios.

The financial database 120 stores financial information collected from the financial information providers 106 and computation results generated by the analytical software 118. The financial information providers 106 is connected to the network 104 via a communication link 126 or the 15 financial information providers may feed the information directly to the host server through a dialup or dedicated line (not shown).

Figure 4 gives a functional view of an implementation of the invention. Structurally, the host server could be implemented as one or more web servers coupled to the network, one or more applications servers running the analytical software and other applications required for the system and 20 one or more database servers that would store the financial database and other information required for the system.

Figure 10 shows an example of a data feed 150 sent from the financial information provider 106 to the host server 102 through the 25 communication link 126. Information is communicated to the host server in the form of messages 151, 152. Each message contains a stream of one or more records 153 each of which carries information about option prices

for an underlying asset. Each message includes header information 154 that identifies the sender and receiver, the current date 155, and an end of message indicator 158, which follows the records contained in the message.

5      Each record 153 in the stream includes an identifier 156 (e.g., the trading symbol) of an underlying asset, an indication 158 of whether the record pertains to a put or call, the strike date 160 of the put or call, the strike price 162 of the put or call, current bid-ask prices 164 of the underlying asset, bid-ask prices 166 for the option, and transaction volumes 168

10     associated with the option. The financial information provider 106 may be an information broker, such as Reuters, Bridge, or Bloomberg, or any other party that has access to or can generate the information carried in the messages. The broker may provide information from sources that include, for example, the New York Stock Exchange and the Chicago Board of

15     Options Exchange.

The financial database 120 stores the information received in the information feed from the financial information providers and other information, including, for example, interest rates and volatilities. The financial database also stores the results generated by the analytical software, including probability distribution functions with respect to the underlying assets and assets that are not the subject of options.

20     The probability distribution information is generated continually (and essentially in real time) from the incoming options data so that the information provided and displayed to users is current. That is, the information is not based on old historical data but rather on current information about option prices.

In addition, other soft information can be accumulated, stored, and provided to users, including fundamental characteristics of the underlying assets, including prices, volatility values, beta, the identification of the industry to which the asset belongs, the yield, the price to book ratio, and the leverage. Other information could include calendars of earnings forecast dates, earnings forecasts, corporate action items, news items that relate to an industry, and the volume of institutional holdings.

5

The messages from the information provider 106 may be sent in response to requests by the host server 102, the information may be sent to the host server 102 automatically at a specified time interval, or the information may be sent as received by the information provider from its sources. The financial database 120 may be maintained on a separate server computer (not shown) that is dedicated to the collection and organization of financial data. The financial database is organized to provide logical 10 relationships among the stored data and to make retrieval of needed information rapid and effective.

10

15

The user 108 may use, for example, a personal computer, a TV set top box, a personal digital assistant (PDA), or a portable phone to communicate with the network 104. Any of these devices may be running an Internet browser to display the graphical user interface (GUI) generated by the host server 102.

20

The host server 102 may provide probability distribution information on the network 104 in the form of web pages and allow the individual user 108, the financial institution 112, the third party web server 114, and the media operator 124 to view the information freely. The host company that runs the host server 102 may generate revenue by, for example, selling 25 advertisement space on its web pages to an advertisement provider 110.

The host server 102 may also provide proprietary information and enhanced services to individual users 108, financial institutions 112, third party web servers 114, and media operators 122 for a subscription fee.

The host server 102 may have a direct link to the financial institutions 112

5 to provide tailored information in a format that can be readily incorporated into the databases of the financial institutions 112. Financial institutions 112 may include, for example, investment banks, stock brokerage firms, mutual fund providers, bank trust departments, investment advisers, and venture capital investment firms. These institutions may incorporate the

10 probability distribution information generated by the analytical software 118 into the financial services that they provide to their own subscribers.

The probability distribution information provided by the host server 102 enables the stock brokerage firms to provide better advice to their customers.

15 A third party web server 114 may incorporate probability distribution

information into its web site. The information may be delivered in the form of an information feed to the third party host of web server 114 either through the Internet or through a dedicated or dial-up connection.

Figure 10 shows an example of a data feed 182 sent from the host server

20 102 to the third party web server 114 through communication link 128.

Data feed 182 carries messages 184 that include header information 186, identifying the sender and receiver, and records 188 that relate to specific underlying assets.

Each record 188 includes an item 190 that identifies a future date, a

25 symbol 192 identifying the asset, risk-neutral probability density

information 193 and cumulative distribution information 194. The record could also include a symbol identifying a second asset 195 with respect to

the identified future date, and so on. Other information could be provided such as a risk premium value with respect to the risk-neutral values.

Examples of third party web servers 114 are the web servers of E\*TRADE, CBS MarketWatch, Fidelity Investments, and The Wall Street Journal. The third party web server 114 specifies a list of assets for which it needs probability distribution information. Host server 102 periodically gathers information from financial information provider 106 and its own financial database 120, generates the probability distribution information for the specified list of assets, and transmits the information to the third party web server 114 for incorporation into its web pages.

Examples of the media operator 124 are cable TV operators and newspaper agencies that provide financial information. For example, a cable TV channel that provides stock price quotes may also provide probability distribution information generated by the host server 102. A cable TV operator may have a database that stores the probability distributions of all the stocks that are listed on the NYSE for a number of months into the future. The host server 102 may periodically send updated information to the database of the cable TV operator. When a subscriber of the cable TV channel views the stock price quotes on a TV, the subscriber may send commands to a server computer to the cable TV operator via modem to specify a particular stock and a particular future date. In response, the server computer of the cable TV operator retrieves the probability distribution information from its database and sends the information to the subscriber via the cable network, e.g., by encoding the probability distribution information in the vertical blank interval of the TV signal.

Likewise, a newspaper agency that provides daily transaction price quotes may also provide the probabilities of stock prices rising above certain percentages of the current asset prices at a predetermined future date, e.g., 6 months. A sample listing on a newspaper may be “AMD 83 88 85 5 ▲40%”, meaning that the AMD stock has a lowest price of \$83, highest price of \$88, a closing price of \$85 that is higher than the previous closing price, and a 40% probability of rising 10% in 6 months.

The analytical software 118 may be written in any computer language such as Java, C, C++, or FORTRAN. The software may include the 10 following modules: (1) input module for preprocessing data received from the financial data sources; (2) computation module for performing the mathematical analyses; (3) user interface module for generating a graphical interface to receive inputs from the user and to display charts and graphs of the computation results; and (4) communications interface 15 module for handling the communications protocols required for accessing the networks.

#### *Web pages and user interfaces*

A variety of web pages and user interfaces can be used to convey the information generated by the techniques described above.

20 For example, referring to figure 5, a GUI 700 enables a user 108 to obtain a range of financial services provided by the host server 102. The user 108 may see the implied probabilities of future prices of marketable assets 706 having symbols 704 and current prices 708. The information displayed could include the probabilities 714 (or 718) of the asset prices rising above 25 a certain specified percentage 712 (or falling below a certain specified percentage 716) of the reference price 710 within a specified period of time 720.

For the convenience of the user 108, GUI 700 includes links 730 to institutions that facilitate trading of the assets. The host company that runs the host server 102 sells advertising space 728 on the GUI 700 to obtain revenue. The GUI 700 also has links 726 to other services provided by the host server 102, including providing advice on lifetime financial management, on-line courses on topics related to trading of marketable assets, research on market conditions related to marketable assets, and management of portfolios of assets.

Referring to figure 6, the GUI 700 also may display an interactive web page to allow the user 108 to view the market's current prediction of future values of portfolios of assets. The past market price 734 and current market price 736 of the asset portfolios 732 are displayed. Also displayed is the price difference 738. The GUI 700 displays the probability 744 (or 746) that the portfolio 732 will gain (or lose) a certain percentage 740 within a specified time period 742. Examples of portfolios include stock portfolios, retirement 401K plans, and individual retirement accounts. Links 748 are provided to allow the user 108 to view the market's current forecast of future price trends of the individual assets within each portfolio.

Referring to figure 7, in another user interface, the GUI 700 displays an interactive web page that includes detailed analyses of past price history and the market's current forecast of the probability distribution of the future values of a marketable asset over a specified period of time. The GUI 700 includes price-spread displays 750 representing the cumulative distribution values of the predicted future prices of an asset over periods of time. The price-spread display 750a shows the price distribution data that was generated at a time three months earlier. A three-month history of the actual asset prices is shown as a line graph for comparison to give the user

108 a measure of the merit of the price distribution information. The price-  
spread display 750b represents the predicted cumulative distribution  
values of the asset prices over a period of one month into the future. The  
left edge of display 750b, of course, begins at the actual price of the asset  
5 as of the end of the prior three-month period, e.g., the current DELL stock  
price of \$50. The probability distribution information implies, for  
example, a 1% probability that the stock price will fall below \$35, and a  
99% probably that the stock price will fall below \$80 in one month. GUI  
700 includes table 752 that shows highlights of asset information and  
graph 754 that shows sector risks of the asset. A box 755 permits a user to  
10 enter a target price and table 757 presents the probability of that price at  
four different future times, based on the calculated implied probability  
distributions.

15 Referring to figure 8, in another approach, a window 402 is displayed on a  
user's screen showing financial information along with two other windows  
408 and 410 showing probability distribution information. The individual  
user 108 could have previously downloaded a client program from the  
host server 102. When the user is viewing any document, e.g., any web  
page (whether of the host server 102 or of another host's server), the user  
20 may highlight a stock symbol 404 using a pointer 406 and type a  
predetermined keystroke (e.g., "ALT-SHIFT-Q") to invoke the client  
program. The client program then sends the stock symbol as highlighted  
by the user to the host server 102. The host server 102 sends probability  
distribution information back to the client program, which in turn displays  
25 the information in separate windows 408 and 410.

When the client program is invoked, a window 422 may be displayed  
showing the different types of price information that can be displayed. In  
the example shown, the "Probability distribution curve" and "Upper/lower

estimate curves" are selected. Window 408 shows the price range of AMD stock above and below a strike price of \$140 from July to December, with 90% probability that the stock price will fall between the upper and lower estimate curves. Window 410 shows the probability density curve  $f(x)$  for 5 AMD stock for a future date of 8/15/2000. The user may also specify a default function curve, such that whenever an asset name is highlighted, the default function curve is displayed without any further instruction from the user.

10 Tabular data such as those shown in TABLE 1 may be generated by the host server 102 and transmitted over the network 104 to devices that have limited capability for displaying graphical data. As an example, the individual user 108 may wish to access asset probability distribution 15 information using a portable phone. The user enters commands using the phone keypad to specify a stock, a price, and a future date. In response, the host server 102 returns the probability of the stock reaching the specified price at the specified future date in tabular format suitable for display on the portable phone screen.

20 Referring to figure 9, a portable phone 500 includes a display screen 502, numeric keys 506, and scrolling keys 504. A user may enter commands using the numeric keys 506. Price information received from the host server 102 is displayed on the display screen 502. Tabular data typically includes a long list of numbers, and the user may use the scroll keys 504 to view different portions of the tabular data.

25 In the example shown in display screen 502, the AMD stock has a current price of \$82. The cumulative distribution values  $F(x)$  for various future prices on 8/15/2000 are listed. The distribution indicates a 40% probability that the stock price will be below \$80 implying a 60% probability of the

stock price being above \$80. Likewise, the distribution indicates an 80% probability that the stock price will be at least \$90, implying a probability of 20% of the stock price being above \$90.

Other embodiments are within the scope of the following claims.

## CLAIMS

- 1       1.     A method comprising:
  - 2           receiving data representing current prices of options on a given
  - 3           asset,
  - 4           deriving from said data an estimate of a corresponding implied
  - 5           probability distribution of the price of said asset at a future time, and
  - 6           making information about said probability distribution available
  - 7           within a time frame that is useful to investors.
- 1       2.     The method of claim 1 in which the data represent a finite number
- 2       of prices of options at spaced-apart strike prices of the asset, and also
- 3       including
  - 4           calculating a set of first differences of said finite number of prices
  - 5           to form an estimate of the cumulative probability distribution of the price
  - 6           of said asset at a future time.
- 1       3.     The method of claim 2 also including
  - 2           calculating a set of second differences of the finite number of
  - 3           strike prices from the set of first differences to form an estimate of the
  - 4           probability distribution function of the price of said asset at a future time.
- 1
- 1       4.     A method comprising:
  - 2           receiving data representing current prices of options on a given
  - 3           asset,

4 deriving from said data an estimate of a corresponding implied  
5 probability distribution of the price of said asset at a future time, and

6 providing a real-time data feed containing information based on  
7 said probability distribution.

1 5. A method comprising:

2 providing a graphical user interface for viewing pages containing  
3 financial information related to an asset; and

4 when a user indicates an asset of interest, displaying probability  
5 information related to the price of the asset at a future time.

1        6.        A method comprising:

2 enabling a user to identify an asset of interest, the asset being one  
3 for which data representing current prices of options on the asset are  
4 available.

5 deriving from said data an estimate of a corresponding implied  
6 probability distribution of the price of said asset at a future time, and

7 providing a display of a probability distribution of prices of the  
8 asset at future times.

## 1 7. A method comprising:

2 enabling a user to indicate a future time and to identify an asset of  
3 interest, the asset being one for which data representing current prices of  
4 options on the asset are available, and

5 displaying to the user a distribution of the probability that the asset  
6 will reach prices within a range of prices at the future time.

1       8.     A method comprising:

2           receiving data representing current prices of options on a given  
3     asset, the options being associated with spaced-apart strike prices of the  
4     asset at a future time,

5           the data including shifted current prices of options resulting from a  
6     shifted underlying price of the asset, the amount by which the asset price  
7     has shifted being different from the amount by which the strike prices are  
8     spaced apart, and

9           deriving from said data an estimate of a quantized implied  
10   probability distribution of the price of said asset at a future time, the  
11   elements of the quantized probability distribution being more finely  
12   spaced than for a probability distribution derived without the shifted  
13   current price data.

1       9.     A method comprising

2           receiving data representing current prices of options on a given  
3     asset, the options being associated with spaced-apart strike prices of the  
4     asset at a future time,

5           deriving from said data an estimate of an implied probability  
6     distribution of the price of said asset at a future time, the mathematical  
7     derivation including a smoothing operation, and

8           making information about said probability distribution available  
9     within a time frame that is useful to investors.

1       10.    The method of claim 9 in which the smoothing operation is  
2     performed in a volatility domain.

1       11.    The method of claim 9 in which the smoothing operation is  
2    performed in the domain of the option prices or in the domain of the  
3    probability distribution information.

1       12.    A method comprising:

2            receiving data representing current prices of options on a given  
3    asset, the options having strike prices at future dates,

4            deriving a volatility for each of the future dates in accordance with  
5    a predetermined option pricing formula that links option prices with strike  
6    prices of the asset;

7            generating a smoothed and extrapolated volatility function;

8            and using the volatility information to generate information within  
9    a time-frame that is useful for investors.

1       13.    The method of claim 12 in which the volatility function is  
2    extrapolated to a wider range of dates than the future dates.

1       14.    The method of claim 12 in which the volatility function is  
2    extrapolated to strike prices other than the strike prices of the options.

1       15.    The method of claim 9 also including

2            generating a smoothed volatility function using only data that are  
3    reliable under a predetermined measure of reliability.

1       16.    The method of claim 9, further comprising:

2            generating an implied volatility function formula having a  
3    quadratic form with two variables representing a strike price and an  
4    expiration date;

5           wherein coefficients of the implied volatility function formula are  
6   determined by applying regression analysis to approximately fit the  
7   implied volatility function formula to each of the implied volatilities.

8   17.   A method comprising:

9           receiving data representing current prices of options on assets  
10   belonging to a portfolio,

11           deriving from said data an estimate of an implied multivariate  
12   distribution of the price of a quantity at a future time that depends on the  
13   assets belonging to the portfolio, and

14           making information about said probability distribution available  
15   within a time frame that is useful to investors.

1   18.   A method comprising:

2           receiving data representing values of a set of factors that influence  
3   a composite value,

4           deriving from said data an estimate of an implied multivariate  
5   distribution of the price of a quantity at a future time that depends on  
6   assets belonging to a portfolio, and

7           making information about said probability distribution available  
8   within a time frame that is useful to investors.

1   19.   The method of claim 18 in which the mathematical derivation  
2   includes generating a multivariate probability distribution function based  
3   on correlations among the factors.

1   20.   A graphical user interface comprising:

2 a user interface element adapted to enable a user to indicate a  
3 future time;

4 a user interface element adapted to show a current price of an  
5 asset; and

6 a user interface element adapted to show the probability  
7 distribution of the price of the asset at the future time.

1 21. A method comprising:

2 continually generating current data that contains probability  
3 distributions of prices of assets at future times,

4 continually feeding the current data to a recipient electronically,  
5 and

6 the recipient using the fed data for services provided to users.

1 22. A method comprising:

2 receiving data representing current prices of options on assets  
3 belonging to a portfolio,

4 receiving data representing current prices of market transactions  
5 associated with a second portfolio of assets, and

6 providing information electronically on the probability that the  
7 second portfolio of assets will reach a first value given the condition that  
8 the first portfolio of assets reaches a specified price at a future time.

1 23. A method comprising:

2 receiving data representative of actual market transactions

3 associated with a first portfolio of assets;  
4 receiving data representative of actual market transactions  
5 associated with a second portfolio of assets;  
6 providing information on the expectation value of the price of first  
7 portfolio of assets given the condition that the second portfolio of assets  
8 reaches a first specified price at a specified future time through a network.

1

1 24. A method comprising  
2 evaluating an event defined by a first multivariate expression that  
3 represents a combination of macroeconomic variables at a time T, and  
4 estimating the probability that a second multivariate expression  
5 that represents a combination of values of assets of a portfolio will have a  
6 value greater than a constant B at time T if the value of the first  
7 multivariate expression is greater than a constant A.

1 25. The method of claim 24 in which the probability is estimated using  
2 Monte Carlo techniques.

1 26. A method comprising  
2 defining a regression expression that relates the value of one  
3 variable representing a combination of macroeconomic variables at time T  
4 to a second variable at time T that represents a combination of assets of a  
5 portfolio, and  
6 estimating the probability that the second variable will have a  
7 value greater than a constant B at time T if the value of the first variable is  
8 greater than a constant A at time T, based on the ratio of the probability of

9 x being greater than A under the regression expression and the probability  
10 of x being greater than A.

1 27. A method comprising

2 defining a current value of an option as a quadratic expression that  
3 depends on the difference between the current price of the option and the  
4 current price of the underlying security, and

5 using Monte Carlo techniques to estimate a probability distribution  
6 of the value at a future time  $T$  of a portfolio that includes the option.

1

## **ABSTRACT**

Data are received that represents current prices of options on a given asset.

An estimate is derived from the data of a corresponding implied probability distribution of the price of the asset at a future time.

- 5 Information about the probability distribution is made available within a time frame that is useful to investors, for example, promptly after the current option price information becomes available.

20092818

10

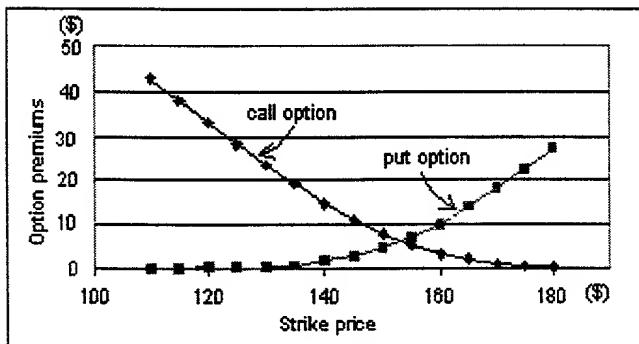


FIGURE 1

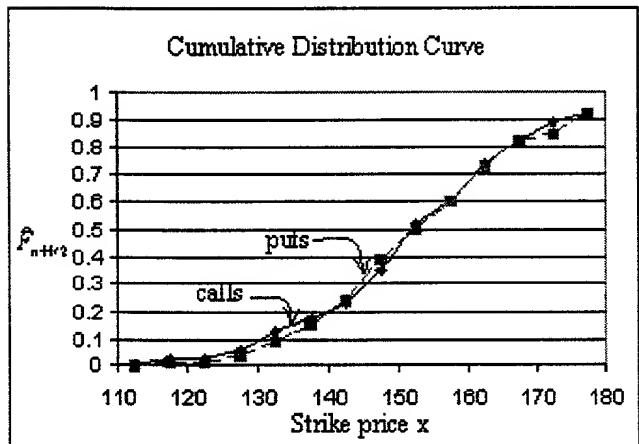


FIGURE 2

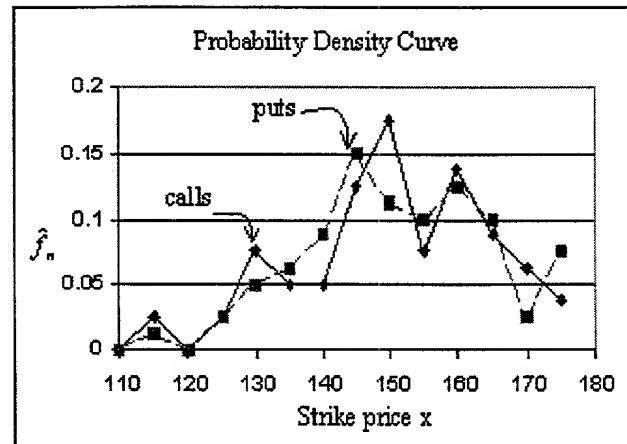


FIGURE 3

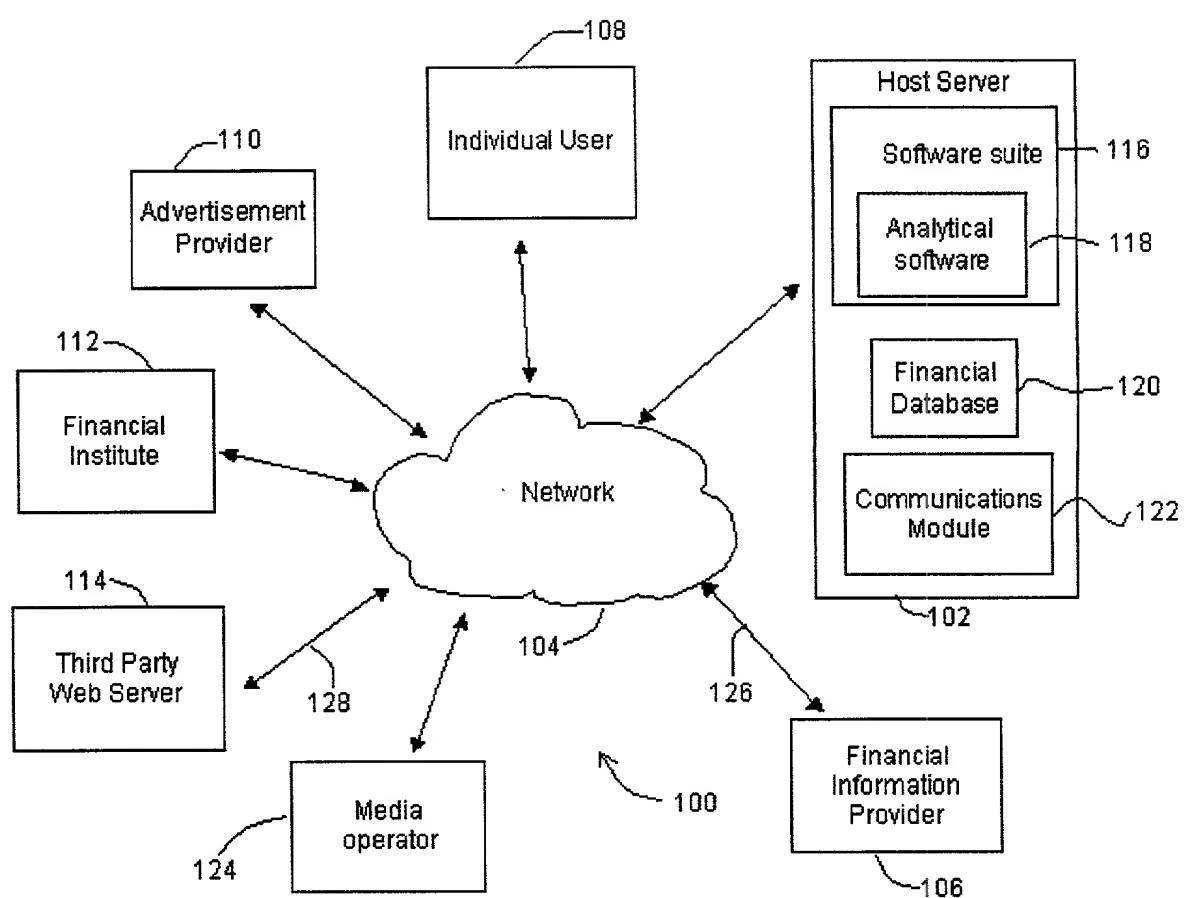


FIGURE 4

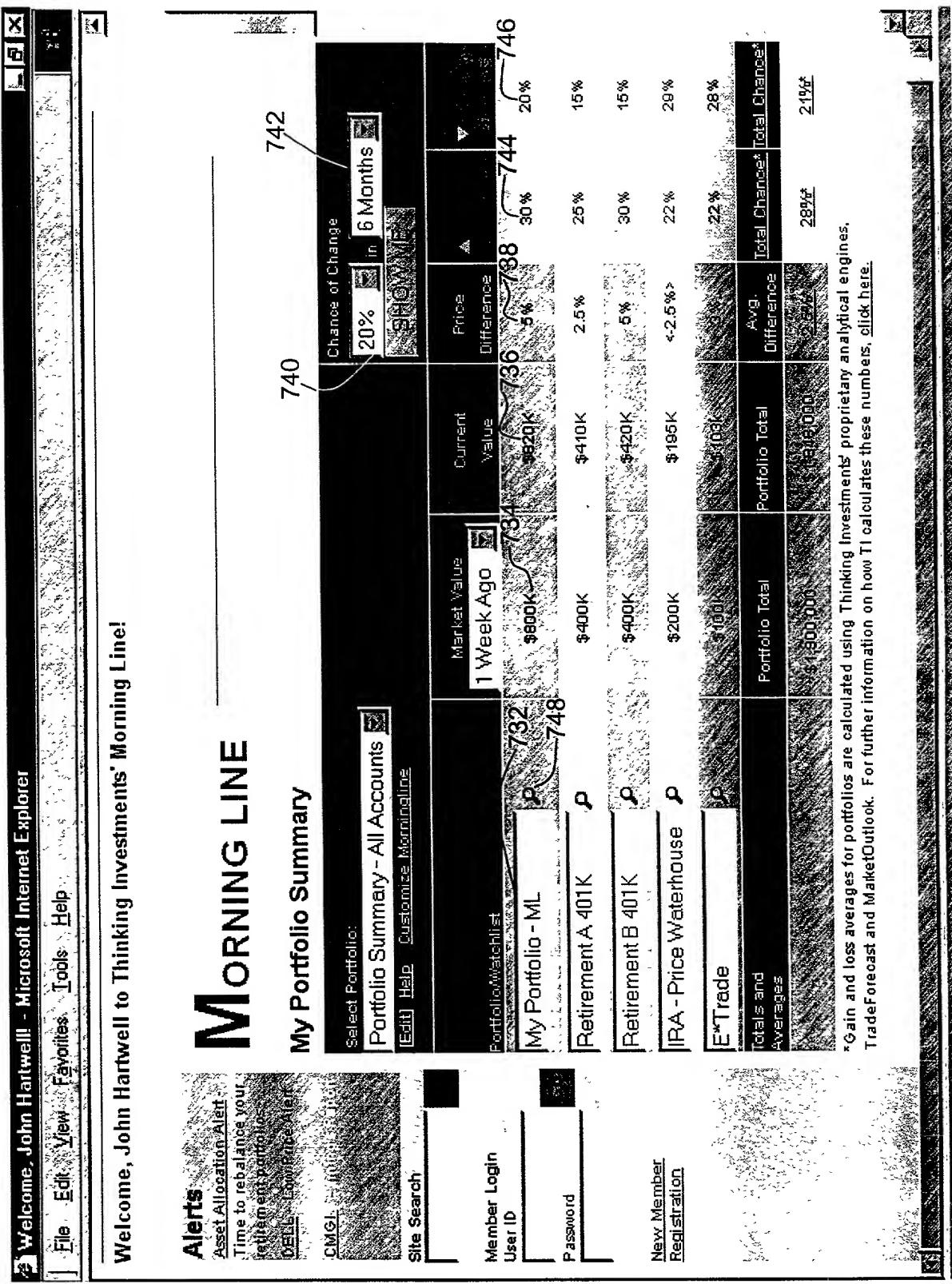


FIGURE 6

http://www.thinkinginvestments.com/rockets/newdemo/morningline.htm - Microsoft Internet Explorer

File Edit View Favorites Tools Help

THINKING INVESTMENTS

Home My Financial Life Advice University Thinking Trader Portfolios Research

Membership About Us Products Help Log In Log Out

726

**Click here to learn more.**

Request a prospectus for more information, including charges and expenses. Read a carefully believed investing. Past performance is no guarantee of future results. Legg Mason Wood Walker, Inc.

**LEGG MASON FUNDS**

ClearFuture™

We can help you reach your retirement goals with tailored 401(k) advice. Click here!

720

710

708

706

704

712

714

716

718

719

720

721

722

723

724

725

726

727

728

**The Morning Line**

Find Symbol

Select Portfolio/WatchList

Today's Morning Line

Historical Results

Select Time Horizon

3 Months

Price Reference

Current

Up %

50

Down %

50

TR

720

710

708

706

704

712

714

716

718

719

720

721

722

723

724

725

726

727

728

729

730

731

732

733

734

735

736

737

738

739

740

741

742

743

744

745

746

747

748

749

750

751

752

753

754

755

756

757

758

759

760

761

762

763

764

765

766

767

768

769

770

771

772

773

774

775

776

777

778

779

780

781

782

783

784

785

786

787

788

789

790

791

792

793

794

795

796

797

798

799

800

801

802

803

804

805

806

807

808

809

810

811

812

813

814

815

816

817

818

819

820

821

822

823

824

825

826

827

828

829

830

831

832

833

834

835

836

837

838

839

840

841

842

843

844

845

846

847

848

849

850

851

852

853

854

855

856

857

858

859

860

861

862

863

864

865

866

867

868

869

870

871

872

873

874

875

876

877

878

879

880

881

882

883

884

885

886

887

888

889

890

891

892

893

894

895

896

897

898

899

900

901

902

903

904

905

906

907

908

909

910

911

912

913

914

915

916

917

918

919

920

921

922

923

924

925

926

927

928

929

930

931

932

933

934

935

936

937

938

939

940

941

942

943

944

945

946

947

948

949

950

951

952

953

954

955

956

957

958

959

960

961

962

963

964

965

966

967

968

969

970

971

972

973

974

975

976

977

978

979

980

981

982

983

984

985

986

987

988

989

990

991

992

993

994

995

996

997

998

999

1000

1001

1002

1003

1004

1005

1006

1007

1008

1009

10010

10011

10012

10013

10014

10015

10016

10017

10018

10019

10020

10021

10022

10023

10024

10025

10026

10027

10028

10029

10030

10031

10032

10033

10034

10035

10036

10037

10038

10039

10040

10041

10042

10043

10044

10045

10046

10047

10048

10049

10050

10051

10052

10053

10054

10055

10056

10057

10058

10059

10060

10061

10062

10063

10064

10065

10066

10067

10068

10069

10070

10071

10072

10073

10074

10075

10076

10077

10078

10079

10080

10081

10082

10083

10084

10085

10086

10087

10088

10089

10090

10091

10092

10093

10094

10095

10096

10097

10098

10099

100100

100101

100102

100103

100104

100105

100106

100107

100108

100109

100110

100111

100112

100113

100114

100115

100116

100117

100118

100119

100120

100121

100122

100123

100124

100125

100126

100127

100128

100129

100130

100131

100132

100133

100134

100135

100136

100137

100138

100139

100140

100141

100142

100143

100144

100145

100146

100147

100148

100149

100150

100151

100152

100153

100154

100155

100156

100157

100158

100159

100160

100161

100162

100163

100164

100165

100166

100167

100168

100169

100170

100171

100172

100173

100174

100175

100176

100177

100178

100179

100180

100181

100182

100183

100184

100185

100186

100187

100188

100189

100190

100191

100192

100193

100194

100195

100196

100197

100198

100199

100200

100201

100202

100203

100204

100205

100206

100207

100208

100209

100210

100211

100212

100213

100214

100215

100216

100217

100218

100219

100220

100221

100222

100223

100224

100225

100226

100227

100228

100229

100230

100231

100232

100233

100234

100235

100236

100237

100238

100239

100240

100241

100242

100243

100244

100245

100246

100247

100248

100249

100250

100251

100252

100253

100254

100255

100256

100257

100258

100259

100260

100261

100262

100263

100264

100265

100266

100267

100268

100269

100270

100271

100272

100273

100274

100275

100276

100277

100278

100279

100280

100281

100282

100283

100284

100285

100286

100287

100288

100289

100290

100291

100292

100293

100294

100295

100296

100297

100298

100299

100300

100301

100302

100303

100304

100305

100306

100307

100308

100309

100310

100311

100312

100313

100314

100315

100316

100317

100318

100319

100320

100321

100322

100323

100324

100325

100326

100327

100328

100329

100330

100331

100332

100333

100334

100335

100336

100337

100338

100339

100340

100341

100342

100343

100344

100345

100346

100347

100348

100349

100350

100351

100352

100353

100354

100355

100356

100357

100358

100359

100360

100361

100362

100363

100364

100365

100366

100367

100368

100369

100370

100371

100372

100373

100374

100375

100376

100377

100378

100379

100380

100381

100382

100383

100384

100385

100386

100387

100388

100389

100390

100391

100392

100393

100394

100395

100396

100397

100398

100399

100400

100401

100402

100403

100404

100405

100406

100407

100408

100409

100410

100411

100412

100413

100414

100415

100416

100417

100418

100419

100420

100421

100422

100423

100424

100425

100426

100427

100428

100429

100430

100431

100432

100433

100434

100435

100436

100437

100438

100439

100440

100441

100442

100443

100444

100445

100446

100447

100448

100449

100450

100451

100452

100453

100454

100455

100456

100457

100458

100459

100460

100461

100462

100463

100464

100465

100466

100467

100468

100469

100470

100471

100472

100473

100474

100475

100476

100477

100478

100479

100480

100481

100482

100483

100484

100485

100486

100487

100488

100489

100490

100491

100492

100493

100494

100495

100496

100497

100498

100499

100500

100501

100502

100503

100504

100505

100506

100507

100508

100509

100510

100511

100512

100513

100514

100515

100516

100517

100518

100519

100520

100521

100522

100523

100524

100525

100526

100527

100528

100529

100530

100531

100532

100533

100534

100535

100536

100537

100538

100539

100540

100541

100542

100543

100544

100545

100546

100547

100548

100549

100550

100551

100552

100553

100554

100555

100556

100557

100558

100559

100560

100561

100562

100563

100564

100565

100566

100567

100568

100569

100570

100571

100572

100573

100574

100575

100576

100577

100578

100579

100580

100581

100582

100583

100584

100585

100586

100587

100588

100589

100590

100591

100592

100593

100594

100595

100596

100597

100598

100599

100600

100601

100602

100603

100604

100605

100606

100607

100608

100609

100610

100611

100612

100613

100614

100615

100616

100617

100618

100619

100620

100621

100622

100623

100624

100625

100626

100627

100628

100629

100630

100631

100632

100633

100634

100635

100636

100637

100638

100639

100640

100641

100642

100643

100644

100645

100646

100647

100648

100649

100650

100651

100652

100653

100654

100655

100656

100657

100658

100659

100660

100661

100662

100663

100664

100665

100666

100667

100668

100669

100670

100671

100672

100673

100674

100675

100676

100677

100678

100679

100680

100681

100682

100683

100684

100685

100686

100687

100688

100689

100690

100691

100692

100693

100694

100695

100696

100697

100698

100699

100700

100701

100702

100703

100704

100705

100706

100707

100708

100709

100710

100711

100712

100713

100714

100715

100716

100717

100718

100719

100720

100721

100722

100723

100724

100725

100726

100727

100728

100729

100730

100731

100732

100733

100734

100735

100736

100737

100738

100739

100740

100741

100742

100743

100744

100745

100746

100747

100748

100749

100750

100751

100752

100753

100754

100755

100756

100757

100758

100759

100760

100761

100762

100763

100764

100765

100766

100767

100768

100769

100770

100771

100772

100773

100774

100775

100776

100777

100778

100779

100780

100781

100782

100783

100784

100785

100786

100787

100788

100789

100790

100791

100792

100793

100794

100795

100796

100797

100798

100799

100800

100801

100802

100803

100804

100805

100806

100807

100808

100809

100810

100811

100812

100813

100814

100815

100816

100817

100818

100819

100820

100821

100822

100823

100824

100825

100826

100827

100828

100829

100830

100831

100832

100833

100834

100835

100836

100837

100838

100839

100840

100841

100842

100843

100844

100845

100846

100847

100848

100849

100850

100851

100852

100853

100854

100855

100856

100857

100858

100859

100860

100861

100862

100863

100864

100865

100866

100867

100868

100869

100870

100871

100872

100873

100874

100875

100876

100877

100878

100879

100880

100881

## FIGURE 5

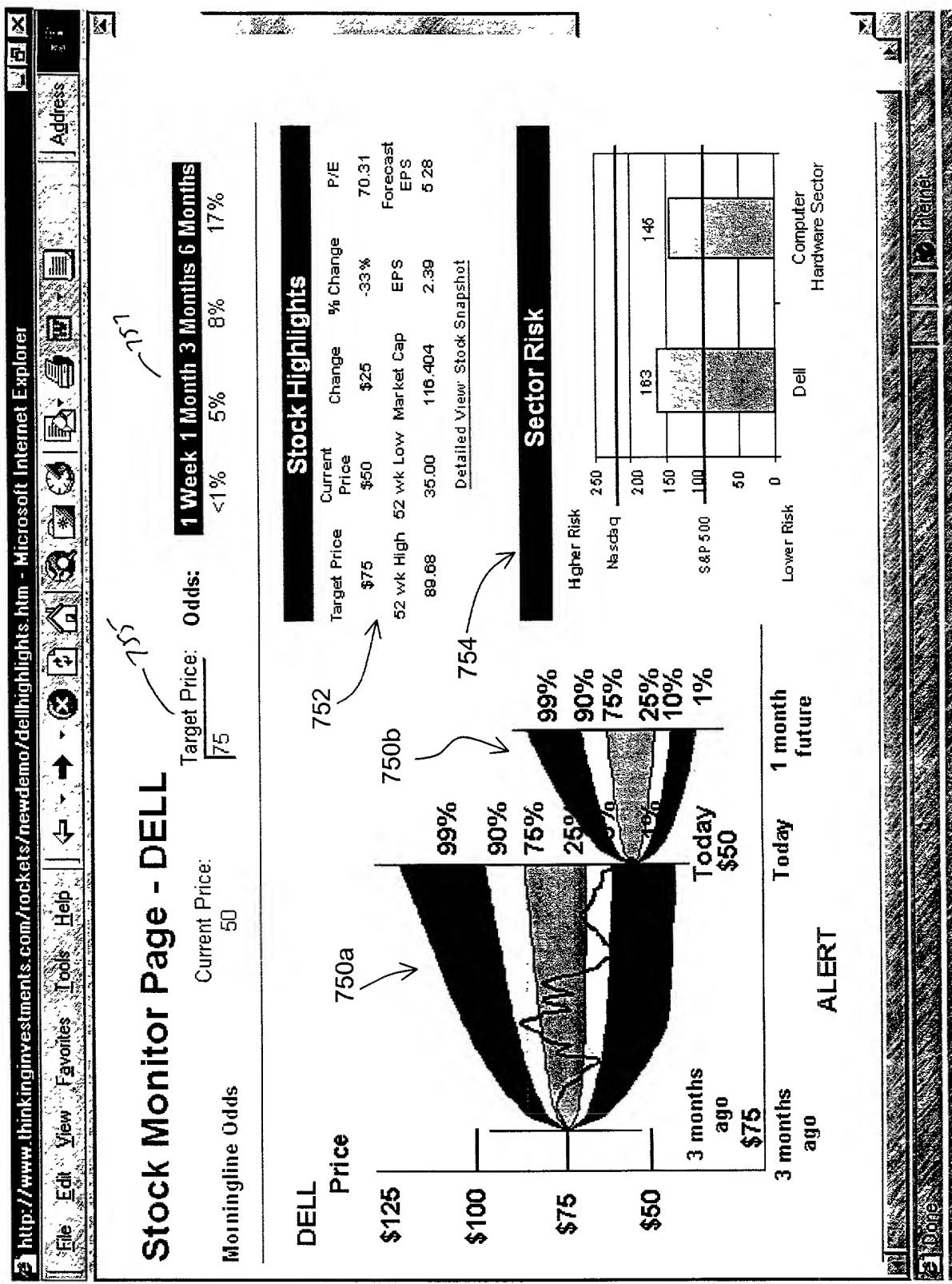


FIGURE 7

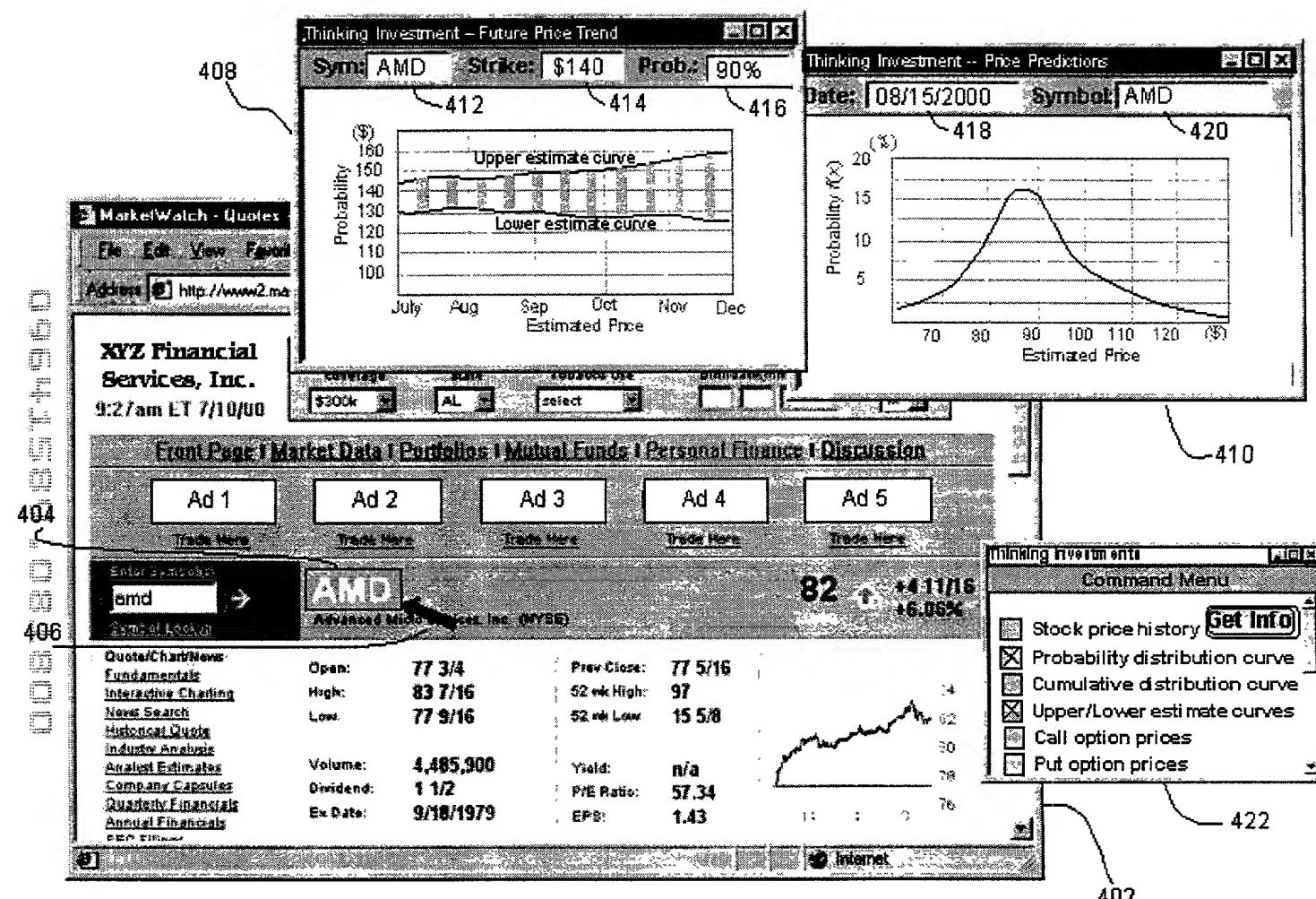


FIGURE 8

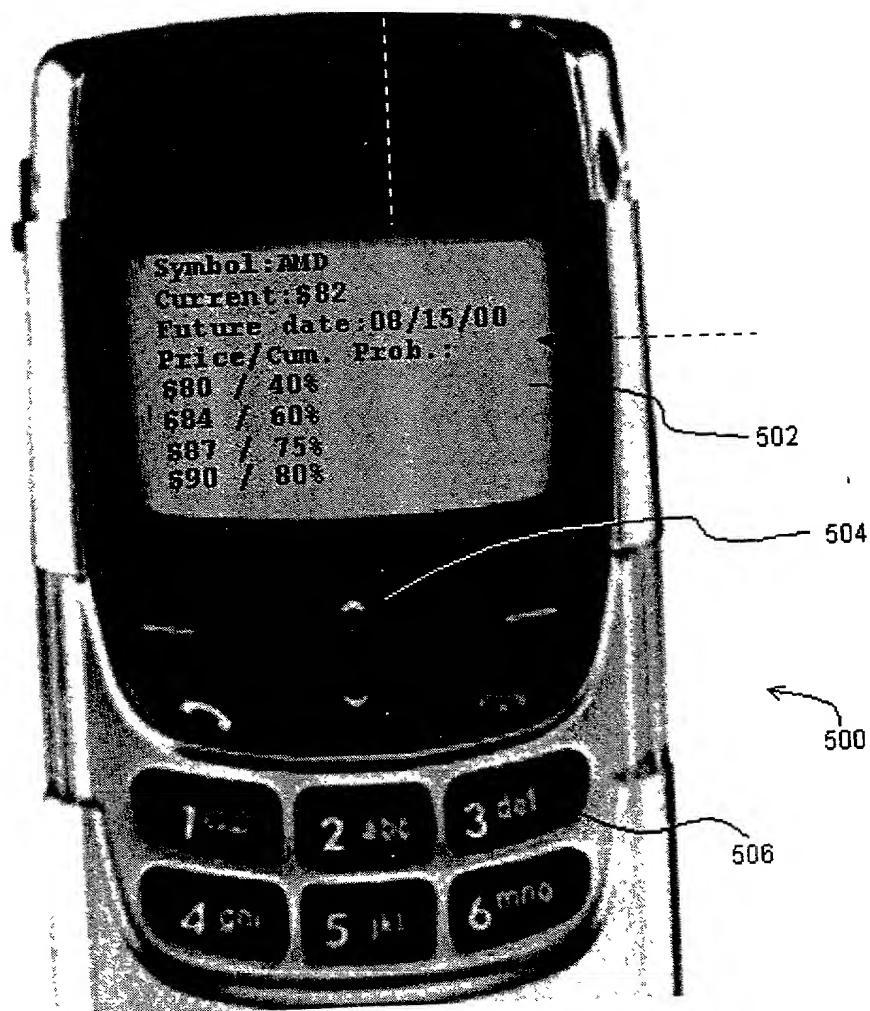


FIGURE 9

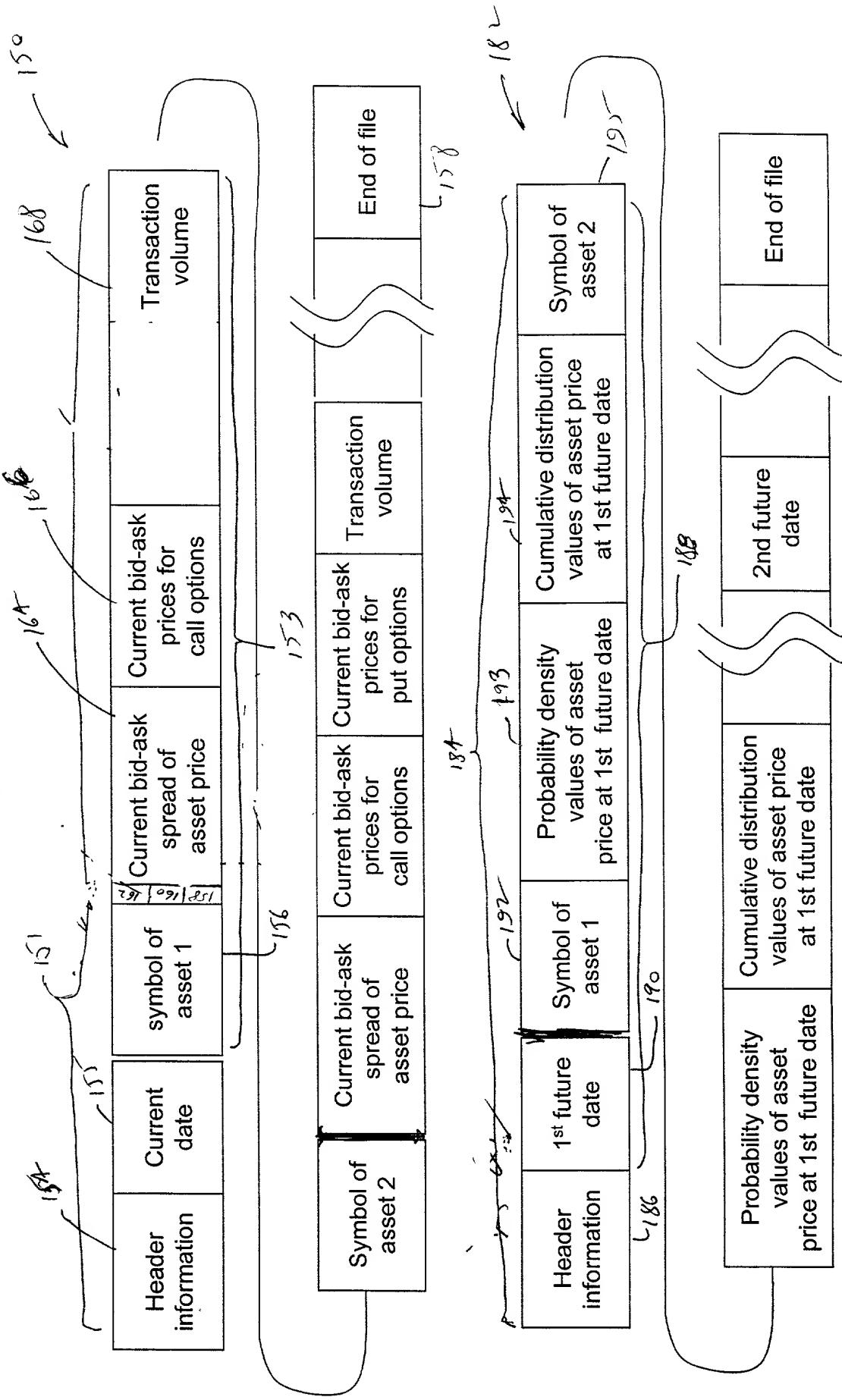


FIGURE 10

## COMBINED DECLARATION AND POWER OF ATTORNEY

As a below named inventor, I hereby declare that:

My residence, post office address and citizenship are as stated below next to my name,

I believe I am the original, first and sole inventor (if only one name is listed below) or an original, first and joint inventor (if plural names are listed below) of the subject matter which is claimed and for which a patent is sought on the invention entitled GENERATING AND PROVIDING INFORMATION ABOUT EXPECTED FUTURE PRICES OF ASSETS, the specification of which:

is attached hereto.  
 was filed on \_\_\_\_\_ as Application Serial No. \_\_\_\_\_ and was amended on \_\_\_\_\_.  
 was described and claimed in PCT International Application No. \_\_\_\_\_ filed on \_\_\_\_\_ and as amended under PCT Article 19 on \_\_\_\_\_.

I hereby state that I have reviewed and understand the contents of the above-identified specification, including the claims, as amended by any amendment referred to above.

I acknowledge the duty to disclose all information I know to be material to patentability in accordance with Title 37, Code of Federal Regulations, §1.56.

I hereby appoint the following attorneys and/or agents to prosecute this application and to transact all business in the Patent and Trademark Office connected therewith:

David L. Feigenbaum, Reg. No. 30,378  
John F. Hayden, Reg. No. 37,460

Robert E. Hillman, Reg. No. 22,837  
Timothy A. French, Reg. No. 30,175

Address all telephone calls to DAVID L. FEIGENBAUM at telephone number (617) 542-5070.

Address all correspondence to DAVID L. FEIGENBAUM at:

FISH & RICHARDSON P.C.  
225 Franklin Street  
Boston, MA 02110-2804

I hereby declare that all statements made herein of my own knowledge are true and that all statements made on information and belief are believed to be true; and further that these statements were made with the knowledge that willful false statements and the like so made are punishable by fine or imprisonment, or both, under Section 1001 of Title 18 of the United States Code and that such willful false statements may jeopardize the validity of the application or any patents issued thereon.

Full Name of Inventor: Philip A. Cooper

Inventor's Signature: \_\_\_\_\_ Date: \_\_\_\_\_  
Residence Address: 61 Buckskin Drive, Weston, MA 02493  
Citizenship: U.S.A.  
Post Office Address: 61 Buckskin Drive, Weston, MA 02493

Full Name of Inventor: Lisette Cooper

Inventor's Signature:

Residence Address: 61 Buckskin Drive, Weston, MA 02493

Date:

Citizenship: U.S.A.

Post Office Address: 61 Buckskin Drive, Weston, MA 02493

Full Name of Inventor: Stewart Myers

Inventor's Signature:

Residence Address: 60 Meriam Road, Lexington, MA 02173

Date:

Citizenship: U.S.A.

Post Office Address: 60 Meriam Road, Lexington, MA 02173

Full Name of Inventor: G. David Forney, Jr.

Inventor's Signature:

Residence Address: 1010 Memorial Drive, Apt. 3G, Cambridge, MA 02138

Date:

Citizenship: U.S.A.

Post Office Address: 1010 Memorial Drive, Apt. 3G, Cambridge, MA 02138

Full Name of Inventor: Leonard L. Scott, Jr.

Inventor's Signature:

Residence Address: 325 Tearose Lane, Charlottesville, VA 22903

Date:

Citizenship: U.S.A.

Post Office Address: 325 Tearose Lane, Charlottesville, VA 22903

**APPENDIX A**

# Appendix A

Three aspects of the invention are:

1. The recognition of the desirability of displaying to a financial investment customer in real time, for example on a World Wide Web site, the probability distribution governing the price of a particular asset (*e.g.*, a stock) at a selected future time.
2. The recognition that such probability distributions can be derived from option prices for that asset, or for related assets, which are readily available in real time.
3. The recognition that probability distributions involving several asset prices simultaneously are useful to investment customers in several contexts, especially in exploring hypothetical scenarios, and that single asset distributions such as (but not restricted to) the above can be meaningfully incorporated into multivariate distributions, manageably determined.

In this appendix we first describe a basic method for deriving probability distributions for single assets from option prices. We next describe improvements on this basic method to address various practical issues. Then we take up the multivariate case and show how to extend this kind of single asset price distribution, or any other, to the multivariate case. Finally, we consider a number of novel multivariate applications, with emphasis on scenario exploration.

## 1 Basic method

A call option is an option to buy an asset (*e.g.*, a stock) at a certain price  $x$  (called the strike price) on a given expiration date  $T$  days in the future. (An option exercisable only on the expiration date is called a European-style option; for simplicity we will consider in this discussion only this type of option.<sup>1</sup>) Similarly, a put option is an option to sell an asset at a strike price  $x$  on a given expiration date. (The “European-style” assumption of no possible early exercise is more important here, but can also be ignored for puts that are not too deeply “in the money.”)

Let  $c(x)$  denote the price of a call option on an asset at strike price  $x$ , and  $p(x)$  the price of a put option. Such prices are established by options market-makers. We have realized that such prices implicitly contain information about a “market view” of the probability distribution of the price of that asset at the expiration date.

In a simple but precise form, this market view can be stated as follows. Suppose that we were given the call price curve  $c(x)$  or the put price curve  $p(x)$  as a continuous function of the strike price  $x$  for all  $x > 0$ . Then, the second derivative of either the call or the put price curve is the market view of the risk-neutral probability density function (pdf)  $f(x)$  of the asset price at the expiration date. In other words,  $f(x) = c''(x) = p''(x)$ .

The idea that option prices determine some kind of implied probability distribution is fairly well known in the financial literature. The idea that a pdf can be computed by taking the second derivative of a continuous option price curve is known in the academic literature, but it does not appear to be very well known. For example, the standard textbook “Options, Futures,

---

<sup>1</sup>Even allowing for possible early exercise, most liquidly traded call options without large dividends can be treated as if there were no possibility of such exercise, since sale of the option is usually a better alternative; therefore, these call options behave similarly to European-style options.

and Other Derivatives,” by John C. Hull (Fourth Edition, 1999; Prentice-Hall) mentions implied probabilities, but not the second-derivative method. The best reference that we have been able to find is J. C. Jackwerth and M. Rubinstein, “Recovering probability distributions from option prices,” *J. Finance*, vol. 51, pp. 1611–1631 (1996), which has only six prior references.

The risk-neutral distribution (at a fixed future time  $T$ , for a fixed asset) is defined as the price distribution that would hold if market participants were neutral to risk, which they generally are not. However, many asset pricing theories, such as those underlying Black-Scholes option theory and most of the variations found in the Hull book above, allow for the true risk-averse asset price distribution to be obtained from the risk-neutral distribution  $f(x)$  just by adjusting the latter by an appropriate risk premium: If there are no dividends, the true distribution is just  $f(xe^{(\mu-r)T})$ , where  $\mu - r$  is the expected annual return rate for the stock in excess of the risk free rate  $r$ . We use a variation on this simple format, slightly modified to allow for dividends (see below), though our invention could also work well with a more complicated adjustment. In this format, a value for  $\mu - r$  must still be supplied. We use as a default the “consensus estimate” taken from the textbook “Active Portfolio Management” (1995) by Grinold and Kahn. These authors note a long-term average value of the risk premium to be 6% per year, and suggest multiplying this number by the stock’s beta to get  $\mu - r$ . The parameter beta is the slope of the line giving a regression of the stock in question against a market portfolio, often taken as the S&P 500. This is the well-known CAPM estimate for the expected excess return. Whether good or bad, its stature as a consensus estimate makes it suited to our aim of providing a market view, though it is only a default. Our invention, which provides the risk-neutral component of the probabilities, could work with other estimates for the risk-averse adjustment parameter  $\mu - r$  and with any explicit scheme for adjusting the risk neutral probability density to the risk-averse probability density. It is worth pointing out that, for shorter time periods—even a month or two—the risk adjustment required is small and generally overwhelmed by fluctuations in the risk-neutral distribution itself.

We give a brief proof that the second derivative procedure gives the correct risk-neutral probability distribution. As in Hull, we may calculate the European call or put price as an expected value in the risk-neutral distribution.

If the actual value of the asset on the expiration date is  $v$ , then the value of a call option at strike price  $x$  is  $\max\{v - x, 0\}$ , and the value of a put option is  $\max\{x - v, 0\}$ . If the actual value is a random variable with pdf  $f(v)$ , then the expected value of a call option at  $x$  at the expiration date is

$$c_T(x) = E_v[\max\{v - x, 0\}] = \int_x^\infty (v - x)f(v) dv,$$

and the expected value of a put option at  $x$  at the expiration date is

$$p_T(x) = E_v[\max\{x - v, 0\}] = \int_0^x (x - v)f(v) dv.$$

The current values  $c(x)$  and  $p(x)$  may be obtained by discounting  $c_T(x)$  and  $p_T(x)$  by  $e^{-rT}$ , where  $r$  is the risk-free interest rate, but for our purposes, forecasting probability distributions at time  $T$ , we do no discounting, and henceforth just write  $c(x) = c_T(x)$ ,  $p(x) = p_T(x)$ .

Parenthetically, from these expressions we observe that

$$p(x) - c(x) = \int_0^\infty (x - v)f(v) dv = x - E_v[v] = x - s^*,$$

where  $s^* = E_v[v]$  is the expected value of the asset at the expiration date under the risk-neutral distribution. (If there are no dividends, then  $s^* = se^{rT}$ ; if there are dividends, then in general it is necessary to subtract from  $se^{rT}$  the value at time  $T$  of the dividends.) This well-known relation is called put-call parity; it shows why either price curve carries the same information.

From the above expression for  $c(x)$ , it follows that its first derivative is

$$c'(x) = - \int_x^\infty f(v) dv = F(x) - 1,$$

where  $F(x) = \int_0^x f(v) dv$  is the cumulative distribution function (cdf) of the random variable  $v$ . To prove this, note that  $v - x = \int_x^v dz$ . Therefore

$$c(x) = \int_x^\infty (v - x) f(v) dv = \int_x^\infty dv \int_x^v dz f(v) = \int_x^\infty dz \int_z^\infty dv f(v) = \int_x^\infty dz (1 - F(z)),$$

where we interchange the variables  $v, z$  to integrate over the two-dimensional region  $\mathcal{R} = \{(v, z) : x \leq z \leq v\}$ . The last expression implies that  $c'(x) = -(1 - F(x))$ .

From put-call parity, it follows similarly that

$$p'(x) = 1 + c'(x) = F(x).$$

Since the cdf and pdf are related by  $F'(x) = f(x)$ , these expressions in turn imply that the second derivative of either  $c(x)$  or  $p(x)$  is the pdf  $f(x)$ :

$$c''(x) = p''(x) = F'(x) = f(x).$$

The general character of the option price curves  $c(x)$  and  $p(x)$  is therefore as follows:

- For all  $x$  less than the minimum possible value of  $v$  (i.e., such that  $F(x) = 0$ ),  $c(x) = E_v[v] - x = s^* - x$  and  $p(x) = 0$ . In other words,  $c(x)$  is a straight line of slope  $-1$  starting at  $c(0) = E_v[v] = s^*$ , while  $p(x) = 0$ .
- For all  $x$  greater than the maximum possible value of  $v$  (i.e., such that  $F(x) = 1$ ),  $c(x) = 0$  and  $p(x) = x - s^*$ . In other words,  $p(x)$  is a straight line of slope  $+1$  and  $x$ -intercept  $s^*$ , while  $c(x) = 0$ .
- These two line segments are joined by a continuous convex  $\cup$  curve whose slope increases from  $-1$  to  $0$  for  $c(x)$ , and from  $0$  to  $+1$  for  $p(x)$ .

We note that the fact that the mean  $E_v[v]$  of the pdf  $f(x)$  is  $s^*$ , the value in future dollars at time  $T$  of the underlying price  $s$  (less the value of any dividends), implies that option prices must be constantly adjusted to reflect changes in the underlying price  $s$ , even if there is no market activity in the options.

The fact that  $s^* = E_v[v]$  also implies that an option price curve can make no prediction about the general direction of the underlying price  $s$ . However, the option price curve does predict the shape of the pdf  $f(x)$ , and in particular its volatility.

## 1.1 Approximations based on a finite subset of bid-asked option prices

In practice, option prices  $c(x)$  and  $p(x)$  are quoted only for a finite subset of equally-spaced strike prices  $x$ , namely  $x = n\Delta$  for integer  $n$  and spacing  $\Delta$ . We denote  $c(n\Delta)$  and  $p(n\Delta)$  by  $c_n$  and  $p_n$ , respectively. Moreover, quotes specify only a bid-asked spread, not exact prices. In this subsection we give methods for dealing with these problems. (Most of the Jackwerth–Rubinstein paper (*op. cit.*) is concerned with these kinds of curve-fitting problems.)

The first derivatives  $c'(x)$  and  $p'(x)$  at  $x = (n + \frac{1}{2})\Delta$  may be estimated by the first differences

$$\hat{c}'_{n+\frac{1}{2}} = \frac{c_{n+1} - c_n}{\Delta}; \quad \hat{p}'_{n+\frac{1}{2}} = \frac{p_{n+1} - p_n}{\Delta}.$$

The corresponding estimates of the cdf  $\hat{F}_{n+\frac{1}{2}} = F((n + \frac{1}{2})\Delta)$  are

$$\hat{F}_{n+\frac{1}{2}} = 1 + \hat{c}'_{n+\frac{1}{2}}; \quad \hat{F}_{n+\frac{1}{2}} = \hat{p}'_{n+\frac{1}{2}}.$$

Thus, using both bid and ask prices for both European-style puts and calls, one can compute four different estimates for the cdf  $\hat{F}_{n+\frac{1}{2}}$ , which can then be combined into a single estimate.

This combination will preferably take into account whether  $x = (n + \frac{1}{2})\Delta$  is much less than the underlying price  $s$  (“deep out-of-the-money”), near  $s$  (“near the money”), or much greater than  $s$  (“deep in-the-money”), according to the different patterns of setting bid-asked spreads in these different ranges. Another consideration is avoiding quotes near prices where early exercise is likely, such as deep in-the-money puts.

Similarly, the second derivatives  $c''(x)$  and  $p''(x)$  at  $x = n\Delta$  may be estimated by the first differences of the estimates of the first derivatives; *e.g.*,

$$\hat{c}''_n = \frac{\hat{c}'_{n+\frac{1}{2}} - \hat{c}'_{n-\frac{1}{2}}}{\Delta} = \frac{c_{n+1} - 2c_n + c_{n-1}}{\Delta^2}.$$

We may take  $\hat{c}''_n$  or  $\hat{p}''_n$  or some combination as above as our estimate  $\hat{f}_n$  of the pdf  $f(n\Delta)$ .

Note that since  $f(x) \geq 0$ , option prices should satisfy a convexity condition, *e.g.*,  $c_{n+1} - 2c_n + c_{n-1} \geq 0$  for call option prices. Indeed, violation of this condition would allow making money via a risk-free “butterfly straddle” involving buying one call option at  $(n + 1)\Delta$  and another at  $(n - 1)\Delta$ , and selling two call options at  $n\Delta$ . A similar result holds for put options.

## 1.2 Dynamic estimates

The methods considered in the previous subsection allow estimation of the cdf and pdf at a subset of  $\Delta$ -spaced values of  $x$ , based on a static set of option quotes at a particular time.

As previously noted, however, option prices must change continually in response to changes in the underlying price  $s$ . Let  $s^*$  denote the corresponding forward price at expiration (the price  $s$  evaluated with interest). Suppose this price (measured in dollars at expiration) moves up (or down) by a small amount, an increment  $\epsilon$  in its logarithm, say, with little or no change in volatility. Here  $\epsilon$  may be viewed as, approximately, the percentage move  $\delta/s^*$  caused by a move of  $\delta$  in the (forward) stock price. We expect in this situation that (forward) probability distribution for the stock price will just be shifted by  $\epsilon$  in the log domain. That is, the distribution will appear to be identical there, except with a mean shifted by  $\epsilon$ . Thus, the value of the new cdf at  $x = e^{\ln x}$  is  $F(e^{(\ln x - \epsilon)}) = F(x/a)$ , where  $F$  denotes the original cdf with distribution mean

$s^*$ , and  $a = e^\epsilon$ . A reasonable call price functional equation that gives the same effect, upon differentiation, is

$$ac(s^*, x/a) = c(as^*, x),$$

where  $c(s^*, x)$  denotes the price, in dollars at expiration, for a call option at strike  $x$  when the underlying is at price  $s$ . Note in this equation that all other variables, such as volatility, are assumed to be the same, which will only be approximately true, even for very small values of  $\epsilon$ .

But, assuming this approximation, we can think of an option price at strike  $x$ , measured when the (forward) price has moved to  $as^*$ , for  $a$  near 1, as giving instead  $a$  times the price of an option at strike  $x/a$ , but corresponding to the current underlying price  $s^*$ . Considering all the strikes at which options are frequently quoted, and thinking additively, we can effectively observe  $c(x)$  (and  $p(x)$ ) for a different subset of approximately equally-spaced strike prices, roughly  $x = n\Delta - \delta$  for various values of  $\delta = \epsilon s^*$ . Some care must, of course, be taken to ensure simultaneity of prices, of option and underlying. For this reason, we may prefer to consider the values of  $n\Delta$  (corresponding to the various standard strike values) separately, and synchronize observed time of sales for an option at a given strike with the underlying security. Implied volatilities (discussed below) could be monitored, to ensure their changes relative to  $\epsilon$  were small.

Using a similar technique to that described in the above paragraphs, meaningful average option prices for a given strike can also be computed, using thin strike intervals and using either short time intervals or time series methods (time averages weighting the present more than the past). Note that, without the framework described in this subsection, the computation of “average” option prices at a given strike are problematic when the stock price varies in the period over which the average is taken.

To summarize: Given enough movements of the underlying price, we can effectively observe prices and compute estimates as above for a much more finely quantized subset of strike prices  $x$ , and provide a framework for improving accuracy through averaging methods.

## 2 Methods for extrapolation and smoothing

There are two major limitations to the basic methods of the previous section. One is that option quotes are available only for certain expiration dates. Another, not so obvious, is that option quotes are reliable primarily for options in which there is substantial market activity. These would typically be nearer-term options at strike prices near the money (the underlying price).

To extend our prediction methods to times other than expiration dates and over wider ranges of strike prices (and also to help reduce “noise” in our displays), we use extrapolation and smoothing techniques. We have found that it is advantageous to do extrapolation and smoothing in the volatility domain.

There are many reasons for this advantage. For example, option practitioners are well aware of the kinds of shapes that the volatility curves (sometimes called “volatility smiles”) have had historically, in various markets, and how these curves vary with time; this can be a guide to imposing structure on the smoothing curves to prevent overfitting of possible artifacts. Many records have been kept of the volatilities implied by option prices, and it is easy to examine how in the past they have changed with respect to price behavior. For example, the Chicago Board of Options Exchange makes public its average near-the-money volatility index (now called VIX) for S&P 100 options back to 1986. Finally, it is easier to work visually with volatility curves, which

would theoretically be flat if  $f(x)$  were lognormal, than with visual differences in near-lognormal pdfs, which can all look very much alike. Mathematically, model improvements can be made in the volatility domain just by changing coefficients of low-degree polynomial approximations, even though these affect higher-order terms in power series for the corresponding cdfs or pdfs.

The following subsections explain more precisely how to work in the volatility domain.

## 2.1 Lognormal pdfs

The standard Black-Scholes theory of option pricing (see Hull, *op. cit.*) yields a lognormal pdf  $f(v)$  whose expected value is  $E_v[v] = s^*$ , such that  $\ln v$  is a Gaussian (normal) random variable with variance  $\sigma^2 T$ , where the parameter  $\sigma$  is called the volatility rate of the asset, and  $T$  is the time to expiration. By a standard property of lognormal distributions, this implies that the mean of  $\ln v$  is  $E_v[\ln v] = \ln s^* - \sigma^2 T/2$ .

From this pdf follows the famous Black-Scholes call option pricing formula [Hull, Appx. 11A]:

$$c(x) = E_v[\max\{v - x, 0\}] = s^* N(d_1(x)) - x N(d_2(x)),$$

where  $N(d_1(x))$  and  $N(d_2(x))$  are values of the cumulative distribution function of a Gaussian random variable of mean zero and variance 1 at the points

$$\begin{aligned} d_1(x) &= \frac{\ln(s^*/x) + \sigma^2 T/2}{\sigma\sqrt{T}} = d_2(x) + \sigma\sqrt{T}; \\ d_2(x) &= \frac{\ln(s^*/x) - \sigma^2 T/2}{\sigma\sqrt{T}} = \frac{E_v[\ln v] - \ln x}{\sigma\sqrt{T}}. \end{aligned}$$

(Recall that our version of the call price is not discounted, and is given in dollars at time  $T$ , and that  $s^*$  is today's stock price, valued in dollars at time  $T$ , less the value of any dividends.) Note that  $\sigma\sqrt{T}$  is the standard deviation of  $\ln v$ ; therefore  $-d_2(x)$  is just  $\ln x$ , measured in standard deviations from the mean  $E_v[\ln v]$ .

Similarly, by put-call parity, we have the Black-Scholes put option pricing formula

$$p(x) = c(x) + x - s^* = s^*(N(d_1(x)) - 1) - x(N(d_2(x)) - 1) = xN(-d_2(x)) - s^*N(-d_1(x)).$$

Taking the derivative with respect to  $x$ , and using  $s^*N'(d_1(x)) = xN'(d_2(x))$  and  $d'_1(x) = d'_2(x)$  (the latter equation holding under the assumption of constant volatility, which we will later drop), we obtain

$$F(x) = c'(x) + 1 = -N(d_2(x)) + 1 = N(-d_2(x)).$$

Now  $F(x)$  is the probability that  $v \leq x$ , which is equal to the probability that  $\ln v \leq \ln x$ , which since  $\ln v$  is Gaussian with mean  $E_v[\ln v]$  and standard deviation  $\sigma\sqrt{T}$  is given by

$$F(x) = \Pr\{v \leq x\} = \Pr\{\ln v \leq \ln x\} = N\left(\frac{\ln x - E_v[\ln v]}{\sigma\sqrt{T}}\right) = N(-d_2(x)).$$

Thus we have verified that the Black-Scholes pricing formulas give the correct cdf  $F(x)$ . The derivative of  $F(x)$  will thus yield the correct lognormal distribution  $f(x) = F'(x)$ .

## 2.2 Characterization of general cdfs

Now let  $F(x)$  be an arbitrary cdf on  $\mathbb{R}_+$ ; *i.e.*, a function that monotonically increases from 0 to 1 as  $x$  goes from 0 to infinity. For simplicity we will assume that  $F(x)$  is strictly monotonically increasing; *i.e.*,  $f(x) = F'(x) > 0$  everywhere. Then there exists a continuous one-to-one “warping function”  $y : \mathbb{R}_+ \rightarrow \mathbb{R}$  such that  $F(x) = N(y(x))$  everywhere; *i.e.*, such that the probability that a random variable  $v$  with cdf  $F(x)$  will satisfy  $v \leq x$  is equal to the probability that a standard Gaussian random variable  $n$  with mean zero and variance 1 will satisfy  $n \leq y(x)$ . Similarly, there is an inverse warping function  $x(y)$  such that  $F(x(y)) = N(y)$ .

Given the warping function  $y(x)$ , the cdf  $F(x)$  may be retrieved from the relation  $F(x) = N(y(x))$ . Therefore the cdf  $F(x)$  completely specifies the warping function  $y(x)$ , and *vice versa*; *i.e.*, both curves carry the same information.

If  $F(x)$  is the cdf of a lognormal variable  $v$  such that  $\ln v$  has mean  $\mathbb{E}_v[\ln v] = \ln s^* - \sigma^2 T/2$  and variance  $\sigma_v^2 = \sigma^2 T$ , as in the previous subsection, then the warping function is given by

$$y(x) = -d_2(x) = \frac{\ln x - (\ln s^* - \sigma^2 T/2)}{\sigma \sqrt{T}} = \frac{\ln x - \mathbb{E}_v[\ln v]}{\sigma_v}.$$

For this reason we may sometimes write  $y(x)$  as  $-d_2(x)$ , even when the cdf is not lognormal so that the right-hand equation above for  $d_2(x)$  does not hold.

## 2.3 Implied volatilities

If  $f(x)$  is not lognormal, then the Black-Scholes pricing formulas do not hold. Nonetheless, given an option price  $c(x)$  or  $p(x)$ , it is common practice to define the implied volatility  $\sigma(x)$  as the value of  $\sigma$  such that the Black-Scholes pricing formula holds, for a given  $x$ ,  $s$  and  $T$ .

The implied volatility curve  $\sigma(x)$  so defined is a function of the strike price  $x$ , which is constant if and only if the pdf  $f(x)$  is actually lognormal. In practice, it is typically a convex  $\cup$  curve, called a “volatility smile.” See, *e.g.*, Hull, Chapter 17.

From Subsection 2.1, we can see that there is a second method of calculating implied volatilities, as follows. Suppose that we have an estimate of the cdf  $F(x)$ . Define the cdf-implied volatility  $\sigma_1(x)$  as the value of  $\sigma$  such that the Black-Scholes cdf formula  $F(x) = N(-d_2(x, \sigma, T))$  holds, for a given  $x$ ,  $s$  and  $T$ .

The first method has the advantages of being defined directly from raw price data, and of being well understood in the financial community. However, the second method has the following advantages:

1. It is easier to calculate, at least from estimates of  $F(x)$ ;
2. It gives a simpler and arguably more intuitive relationship between volatility and the cdf  $F(x)$ . If we use the traditional implied volatility  $\sigma(x)$ , then the relationship is instead

$$F(x) = N(-d_2(x)) + \frac{\partial c}{\partial \sigma} \sigma'(x).$$

3. It fits better with the multivariate theory to be developed below.

We have observed that the two curves  $\sigma(x)$  and  $\sigma_1(x)$  seem to be fairly similar, at least as to the direction of their slope, and are generally not too far apart in value “near the money”. Also  $\sigma_1(x) = \sigma(x)$  whenever  $\sigma(x)$  has zero slope, though  $\sigma_1(x)$  is a little smaller than  $\sigma(x)$  when the slope  $\sigma'(x)$  is negative (which often occurs for stocks). See the above equation. Finally, one function is as *ad hoc* as the other. Therefore, because of the above reasons, we generally prefer to use the cdf-implied volatility curve  $\sigma_1(x)$ .

In any case, it is clear that either  $\sigma(x)$  or  $\sigma_1(x)$  contains the same information as any of the curves  $c(x)$ ,  $p(x)$ ,  $F(x)$  or  $f(x)$ . From  $\sigma(x)$  or  $\sigma_1(x)$  we can recover  $c(x)$  or  $F(x)$  using the Black-Scholes call option pricing or cdf formula, and from this we can obtain all other curves.

## 2.4 Extrapolation and smoothing in the volatility domain

The volatility curve  $\sigma(x)$  or  $\sigma_1(x)$  may be calculated pointwise from the corresponding curve  $c(x)$  or  $p(x)$  to give a set of values at a finite subset of strike prices  $x$ . Each of these values may be deemed to have a certain degree of reliability.

It is then a standard problem to fit a smoothed and extrapolated curve  $\tilde{\sigma}(x)$  or  $\tilde{\sigma}_1(x)$  to these points, taking into account their relative reliabilities. Any standard smoothing and extrapolation method may be used. In general, the usual problems of avoiding overfitting or oversmoothing must be addressed.

It is well-known that implied volatilities also vary with time. We generally wish to estimate curves  $\tilde{\sigma}(x, T)$  or  $\tilde{\sigma}_1(x, T)$  as replacements for the constant volatility  $\sigma$  in the Black-Scholes formulas, *e.g.*,  $c(x) = c(x, \sigma, T)$  or  $F(x) = N(-d_2(x, \sigma, T))$ .

In an especially meaningful example, we have experimented with a class of smoothing algorithms used in “Implied volatility functions: Empirical tests,” by B. Dumas, J. Fleming and R. E. Whaley, *J.Finance*, vol. 53, pp. 2059-2106, Dec. 1998. These authors fit an implied volatility curve  $\sigma(x)$ , for the purpose of setting up a “strawman” option price model for testing (and defeating) a theory regarding the role of volatility in option pricing. Their “strawman” option pricing model  $c(x)$  was obtained by putting the resulting smoothed curve back into the Black-Scholes call formula. It is a “strawman” *ad hoc* model, because no intuitive notion of stock volatility could possibly vary with strike price, which the stock never “sees.” Nevertheless, their model performed admirably, surpassing in predictive power the highly regarded “implied tree” method. One possible explanation offered was that their model mimicked in a smooth way interpolation methods actually employed by practitioners in the options markets. (See the discussion of “Volatility matrices” in Hull, cited above.) Such an approach to option pricing seems ideal to us, because of its accuracy and because its underlying rationale represents a market view. Thus, we use the Dumas-Fleming-Whaley model for our own entirely different purpose, that of forecasting probability distributions. All that is necessary is to differentiate their call price model, which, conveniently for us, is a smooth function of strike price and other standard variables such as time, current stock price, and the risk-free rate of interest. The formula for the cdf  $F(x)$  is, as before, this derivative with 1 subtracted, or

$$F(x) = N(-d_2(x)) + \frac{\partial c}{\partial \sigma} \sigma'(x).$$

We can make this very explicit. We have

$$\frac{\partial c}{\partial \sigma} = x\sqrt{T}N'(-d_2(x))$$

where  $N'(z)$  denotes standard normal density, while  $\sigma'(x)$  may be computed by differentiating the Dumas-Fleming-Whaley fitted volatility curve. The latter has the form

$$\sigma(x, T) = a_0 + a_1 x + a_2 x^2 + a_3 T + a_4 T^2 + a_5 x T.$$

The coefficients  $\{a_i\}$  are determined by regression. This kind of quadratic curve-fitting is easily implemented. Dumas-Fleming-Whaley impose a constraint to prevent their volatilities from going below 0 (or even below 0.01), and we have imposed further constraints on extrapolations (which we often carry out beyond the range of their tests), to ensure that the final cdf does not go below zero or above one. We have experimented with other variations on their basic approach, for example, using linear interpolation in the time domain, where we do not need to take derivatives. Our methods would, of course work, with any approach, possibly quite different, to volatility curve-fitting, though the general Dumas-Fleming-Whaley approach has many things going for it: accuracy, conformity to marketplace use of Black-Scholes, smoothness (differentiability, in particular), conformity to historical experience regarding the smile structure of volatility curves (especially important for extrapolation), and simplicity (which, beyond ease of implementation, helps avoid overfitting). These advantages are achieved in a probability context that was not considered in the paper where these volatility curves were introduced.

### 3 The multivariate case

The methods in the previous sections are capable of generating a display of raw or smoothed and extrapolated probability distributions for any optionable asset. Option prices are quoted on a large number of securities, as well as on certain indices, such as the S&P 500.

However, an investor would also like to know future probability distributions for:

- His or her entire portfolio;
- Mutual funds;
- A security without a quoted option;
- A security in a hypothetical scenario.

All of these questions involve considerations of several securities at once, and the probabilities of their simultaneous configuration of prices. This is clearly a consideration in the first two items above, but also enters in the third, where we would want to extract as much information as possible about the security without a quoted option price from those correlated with it that do have quoted options. Finally, in scenario analysis there are many questions that involve considering the probabilities of several security prices occurring at once, including changes in factors influencing the market that might be modeled by changes in a portfolio of those securities most affected. We will take all of these issues up in the remainder of this document, but for now we just try to give a basic introduction.

For a portfolio of securities, or a mutual fund, we are interested in a composite asset of the form

$$x = h_1 x_1 + h_2 x_2 + \cdots + h_n x_n,$$

where the  $x_i$  are all assets for which we individually know the cdf  $F(x_i)$  or the pdf  $f(x_i)$ . To give our method the most flexibility, we do not require that this knowledge come from any particular

procedure, though we favor the approach of the preceding two sections. However, even for some securities or indices with a quoted option, we might not feel there was sufficient option activity to justify a full fit of a volatility curve, and might take a cruder substitute, even a flat straight line based on an average of available implied volatilities. In addition, it is convenient to allow the possibility that a few assets we are monitoring might not have any quoted option at all; this is easily accommodated by, say, using a flat volatility curve with a historical value for volatility. For testing purposes and comparisons we might even want to consider a list of assets with all volatility curves given this way. In any case our methodology here is very general, and we only require that we know warping functions  $y_i(x_i)$  such that  $F(x_i) = N(y_i(x_i))$  for all  $i$ . If the asset has an active options market, then the warping functions may be determined by either first estimating  $F(x_i)$  directly from (finite differences of) options price data, as in subsection 2.2, or by using the approach discussed later in Section 2 of extrapolating and smoothing in the volatility domain. In the latter case we have an explicit form of the warping function  $y_i(x_i)$  in terms of a fitted volatility curve  $\tilde{\sigma}_1(x_i, T)$  as  $y_i(x_i) = -d_2(x_i, \tilde{\sigma}_1(x_i, T), T)$ , and this equation can also be used with any volatility curve with the assets above, that might have fewer or no traded options. In a later section we will discuss portfolios in the logarithm domain, possibly containing long and short positions. One can think of warping them to standard normal directly, subtracting the mean and dividing by the standard deviation. Alternately, to keep our notation uniform, one can invent an asset with price  $x_i$  such that  $-d_2(x_i)$  gives this warped value (using for  $\sigma_1(x)$  the observed historical volatility). But we wish to emphasize that the method we are describing works with ANY single-variable warping functions, even using a different one for each variable. The only further substantitive ingredient is the plausibility of using JOINTLY normal distributions, which we now discuss.

The general problem is to find a multivariate probability distribution for the complete set of variables  $(x_1, \dots, x_n)$ , or equivalently for their logarithms. In simple financial models generalizing the Black-Scholes framework, the multivariate distribution of the logvariables is multivariate (*i.e.*, jointly) normal; see Musiela and Rutkowski's book "Martingale methods in financial markets" (1999). This implies that all portfolios of these logvariables are jointly normal, and can also be used with other logvariables and portfolios of them to form a jointly normal distribution. Thus, if we wish, it is reasonable to use BARRA (or functionally equivalent) factors as single (log)variables in our model, using, say, individual normal distributions for them based on historical volatility. These factors may represent fundamentals of companies or even macroeconomic variables such as interest rates. We do not further discuss such factors, but refer to the book of Grinold and Kahn cited above, which also describes how to closely approximate them as portfolios of security returns. Our preference is to not use BARRA factors directly, but stay as much as possible in the world of optionable securities, and address questions involving BARRA factors in terms of approximating portfolios consisting mostly of optionable securities. (But for testing and comparisons, it is still useful to be able to include them directly, and we do have that capability.)

Now we certainly do not wish to use only the simple multidimensional Black-Scholes model, which would not directly allow the nonlognormal input from our single-variable distributions based on the options markets. At the same time, option prices on individual assets do not tell us anything about how assets interact, in particular, their correlations. Fortunately, correlations may be estimated from past (historical) data, and may be viewed as covariances for data that has been standardized (has standard deviation 1). Each multivariate normal distribution is determined by its mean and covariance matrix. Thus, a natural approach is to use the individual distributions to transform or "warp" the variables to standard normal, then impose a

multivariate normal structure based on the correlation matrix. This procedure is independent of the individual warping functions, which may be different for different individual variables, and in particular, can incorporate our market-based option distributions for individual variables representing securities with active options markets. A slightly different approach is to use correlations of the warped variables. This procedure is likely to be more accurate, but may involve more computational time.

We indicate some details. As before, it is notationally convenient to use  $v_i$  as a second notation for  $x_i$ , favoring the latter for fixed values and the former as a variable. Let  $C$  be the historical correlation matrix of the log variables  $(\ln v_1, \dots, \ln v_n)$ , whose entries are the cross-correlations

$$\rho_{ij} = \frac{\mathbb{E}(\ln v_i \ln v_j) - \mathbb{E}(\ln v_i)\mathbb{E}(\ln v_j)}{\sqrt{\mathbb{E}(\ln v_i)^2}\sqrt{\mathbb{E}(\ln v_j)^2}}.$$

Then all diagonal terms  $\rho_{ii}$  are equal to 1, and  $C$  is a positive semi-definite covariance matrix, which we may here assume to be nonsingular (positive definite). If we use instead correlations of warped variables, we have simply

$$\rho_{ij} = \mathbb{E}(y_i y_j).$$

Let us define  $F_C(y_1, \dots, y_n)$  as the cdf of a multivariate Gaussian random variable with mean zero and covariance matrix  $C$ . Thus,  $F_C(b_1, \dots, b_n)$  is the probability that each variable  $y_i$  is at most some value  $b_i$ . There are more elaborate versions, such as  $F_C(a_1, \dots, a_n; b_1, \dots, b_n)$ , giving the probability that each  $y_i$  satisfies  $a_i \leq y_i \leq b_i$ . In the single-variable case these latter functions are obtained from the simple cdf by a single subtraction, involving two terms, but the corresponding bivariate case involves four terms, and in  $n$  dimensions there would be  $2^n$  terms. However, each of these more elaborate cdf's can be directly computed as an integral, just like the simple cdf. Since the more elaborate cdf's are needed for Monte Carlo calculations, possibly in high dimensions, it is best to think of them as being computed directly.

We then define the multivariate cdf's

$$F(x_1, \dots, x_n) = F_C(y_1(x_1), \dots, y_n(x_n)),$$

and

$$F(a_1, \dots, a_n; b_1, \dots, b_n) = F_C(y_1(a_1), \dots, y_n(a_n); y_1(b_1), \dots, y_n(b_n))$$

where the  $y_i(x_i)$  are the known warping functions for the individual variables. We find it convenient, with some abuse of language, to speak of  $F(x_1, \dots, x_n)$  as "the cdf", even though we have all of the above functions in mind, and to use  $F(x_1, \dots, x_n)$  as a proxy for the whole distribution (which it does, theoretically, determine). This multivariate cdf then has the following properties:

- Since the marginals of  $F_C(z_1, \dots, z_n)$  are Gaussian with mean 0 and variance 1, the marginals of  $F(x_1, \dots, x_n)$  are equal to  $N(y_i(x_i)) = F(x_i)$ ; i.e., they are correct according to each single-variable model.
- If the logvariables  $(\ln v_1, \dots, \ln v_n)$  are actually jointly Gaussian, then the multivariate cdf  $F(x_1, \dots, x_n)$  is correct.

In summary, the true joint distribution is approximated by a jointly lognormal distribution using historical correlations, combined with warping functions on each variable such that the marginal distribution of each variable is correct according to a selected single-variable model

(for example, according to our single-variable model for optionable securities, or according to the lognormal model using historical volatility). The single variables may actually be portfolios, with a default distribution for the portfolio return being lognormal, based on historical volatility. This multivariate theory generalizes both our single-variable theory and standard multivariate (log)Gaussian models. It again allows for market input through option prices, to the extent that components have an active option market, but does not exclude nonoptionable securities, and also allows portfolios as single variables. In this way BARRA (or functionally equivalent) factors are also allowed because of their interpretation as portfolios of long and short positions.

## 4 Applications to portfolios

Given the multivariate cdf  $F(x_1, \dots, x_n) = F_C(y_1(x_1), \dots, y_n(x_n))$ , we can answer many typical questions. We first give an overview, and then take up some of the applications in more detail.

As one example, suppose that we want to find the cdf of a portfolio variable

$$x = h_1 x_1 + h_2 x_2 + \dots + h_n x_n,$$

where the  $h_i$  are arbitrary coefficients. A simple Monte Carlo method, probably not the fastest, is to draw random samples from the jointly Gaussian distribution with cdf  $F_C(y_1, \dots, y_n)$ , transform each  $y_i$  via the inverse mapping function  $x_i(y_i)$ , and then compute the resulting output sample

$$x = h_1 x_1(y_1) + \dots + h_n x_n(y_n).$$

After enough samples, we will have an approximation to the cdf of  $x$ . More precisely, the probability that  $a \leq x \leq b$  is, approximately, the average number of samples  $y_1, \dots, y_n$  with  $a \leq h_1 x_1(y_1) + \dots + h_n x_n(y_n) \leq b$ , and this approximation becomes exact in the limit for large sample sizes. This works for real portfolios, or for portfolios constructed from a number of assets and a residual variable, as might arise from a regression. Usually the regression is done in the log domain, which we discuss below. Note that the Monte Carlo method just described works perfectly well if the expression for  $x$  above is replaced by any function  $f(x_1, \dots, x_n)$  of the  $x_i$ , possibly quite nonlinear.

### 4.1 Log domain portfolios

In this subsection, we point out how our methods fit with another paradigm in common use in the financial community, and set up some further notation. It is common to work in the return domain, or equivalently, with logarithms; *i.e.*,

$$\ln x = \beta_1 \ln x_1 + \dots + \beta_n \ln x_n.$$

Ignoring any possible identification of these variables with those in the previous section, the same discussion and Monte Carlo method as above applies, if we regard  $x$  as a nonlinear portfolio  $x = f(x_1, \dots, x_n) = \exp(\beta_1 \ln x_1 + \dots + \beta_n \ln x_n)$ . If the sum  $B$  of the  $\beta_i$ 's is 1, such an  $x$  may be written  $x = \hat{h}_1 x_1 + \hat{h}_2 x_2 + \dots + \hat{h}_n x_n$  where  $\hat{h}_i = \beta_i x / x_i$ . Even if  $B$  is not 1, incremental changes (“returns”)  $d \ln x$  computed from this equation for  $x$  are consistent with the above expression for  $\ln x$ . It is common in the financial community to think of  $\hat{h}_i$  as approximately a constant  $h_i$ , so

that for short periods, where the  $x'_i$ 's do not change too much, this equation for  $x$  is comparable to the portfolio equation in the previous subsection.<sup>2</sup>

For an asset  $x$  not given explicitly in terms of the terms of the  $x_i$ , we obtain a similar expression via linear regression:

$$\ln x = \beta_0 \ln x_0 + \beta_1 \ln x_1 + \cdots + \beta_n \ln x_n,$$

The  $\beta_i$  for  $i \neq 0$  are correlation coefficients chosen to minimize the variance of the residual in historical data (perhaps subject to constraints, such as  $\beta_i \geq 0$  and  $\sum_{i=1}^n \beta_i = 1$ ). For example,  $x$  might be a security without a quoted option, and the  $x_i$  for  $i \neq 0$  could be taken as assets for which we individually know the probability distributions, in addition to the required correlation coefficients for  $x$ . We have written the residual term as  $\beta_0 \ln x_0$  (usually thinking of  $\beta_0 = 1$  and the residual as normally distributed).<sup>3</sup> The mean of the latter could be nonzero, giving the regression “alpha”—a constant term making the mean of the regression correct. Alternatively, we could modify the equation to allow an explicit alpha, and keep the residual mean zero. Another minor variation might include the addition of a dummy variable with constant return, to adjust the value of  $x$  up or down. In particular, this gives another way of adjusting the residual mean to zero. This equation gives the previous one as a special case if we allow  $\beta_0 = 0$ .

#### 4.1.1 Fast fits of portfolios

One approach, which promises to be relatively fast computationally, is the following. As in the development of cdf-implied volatilities in Section 2, let us assume that each logvariable  $\ln x_i$  above is “Gaussian” with nonconstant variance  $\sigma_1(x_i)^2 T$ . In other words, the cdf is given by  $F(x_i) = N(-d_2(x_i, \sigma_1(x_i), T))$ . Our aim will be to give  $F(x)$  by a similar equation, using some kind of fitted curve  $\sigma_1(x)$ . We will assume that we have some class of volatility curves in mind, with a small number of parameters which must be determined.

If the variables  $\ln x_0, \dots, \ln x_n$  were truly jointly Gaussian, then  $\ln x$  would also be Gaussian. Its variance would be given by the formula

$$\text{Var}(\ln x) = \sum_{i,j} \beta_i \sigma_i \rho_{ij} \beta_j \sigma_j T,$$

where  $\rho_{ij}$  is the correlation between  $\ln x_i$  and  $\ln x_j$ , and  $\sigma_i^2 T = \text{Var}(\ln x_i)$ . We therefore define the estimate  $\hat{\sigma}_1(x)$  of  $\sigma_1(x)$  by the conditional expectation

$$\hat{\sigma}_1(x) = E(\beta_i \sigma_1(x_i) \rho_{ij} \beta_j \sigma_1(x_j) \mid \ln x = \sum_i \beta_i \ln x_i).$$

The calculation of the above conditional expectation may be done with Monte Carlo methods. In the language of nonlinear portfolios above, we would take the function  $f(x_1, \dots, x_n)$  to be 0 outside a thin multidimensional solid enclosing the hyperplane defined by  $\ln v = \sum_i \beta_i \ln v_i$ . Inside the solid we would take  $f(x_1, \dots, x_n)$  equal to the above expression for  $\text{Var}(\ln x)$ , divided by the probability of being in the solid (also a Monte Carlo calculation). In terms of samples, we just take the average of  $\text{Var}(\ln x)$  over all the samples that end up inside the thin solid. However,

<sup>2</sup>Thus  $\hat{h}_i \frac{dx_i}{x} = \beta_i \frac{dx_i}{x_i} = \beta_i d \ln x_i$ , so that for small changes  $dx$ , the change  $dx$  from the first equation is approximately the same as would be obtained from the second. However, this relationship requires “rebalancing” to remain a good approximation for longer periods.

<sup>3</sup>For the residual term  $i = 0$ , we can use a constant variance, or impose some generic nonconstant structure based on observed behavior.

it is not necessary to compute all values of  $\hat{\sigma}_1(x)$ , but only enough to fit the parameters for the volatility curves we are using.

The estimated mean of  $\ln x$  would be  $\ln s^* - \frac{1}{2}\hat{\sigma}_1(x)^2$ , with  $s^*$  determined as before, or replaced with some risk-averse estimate, to obtain the risk-averse or “true” distribution. (It is common, incidentally, to use factor models such as these to estimate a risk-averse version of  $\ln s^* = \sum_i \beta_i \ln s_i^*$  from risk-averse values of  $s_i^*$ .)

Also, we mention here one useful variation: We may prefer not to view the residual term  $\beta_0 \ln x_0$  as part of the model, and instead write down a joint pdf only for  $\ln x_1, \dots, \ln x_n$ . In this case we can use the double expectation

$$\hat{\sigma}_1(x) = E(E(\sum_{i,j} \beta_i \sigma_1(x_i) \rho_{ij} \beta_j \sigma_1(x_j) \mid \ln x = \sum_i \beta_i \ln x_i)),$$

where the inner expectation is with respect to the variables  $x_1, x_2, \dots, x_n$ , and the outer expectation is with respect to the residual. We might take the standard deviation  $\sigma(x_0)$  of the residual (taking  $\beta_0 = 1$ ) as a constant, determined historically, or make an estimate based on some leverage model.

Now we can estimate the cdf  $F(x)$  by

$$\hat{F}(x) = N(-d_2(x, \hat{\sigma}_1(x), T))$$

as in the univariate case. To summarize, we use our multivariate model to determine parameters for a univariate model of the portfolio. After that is done, we can obtain probabilities for the portfolio without having to go back to the multivariate model, thus achieving a savings in time. We could take this one step further and think of randomly generating values of  $\sigma_1(x, T)$  independently of any Monte Carlo philosophy (but perhaps still throwing away values of  $x$  too far out-of-the-money), and then using the values obtained to do the regression required in the Dumas-Fleming-Whaley approach.

## 5 “What-if” questions

The multivariate distribution lends itself to the study of many questions regarding conditional probabilities. For example, suppose that we want to know the effect of the increase or decrease of some segment of the market on a portfolio, or the increase or decrease of some macro-economic factor. BARRA, following earlier ideas of Ross, has viewed such macro-economic factors as portfolios with both long and short positions. Similarly, BARRA considers market segments associated to price-to-earnings ratios and other fundamental parameters, as well as to industry groupings, as portfolios. (See the book of Grinold-Kahn cited above.) Thus, we are led simply to consider the effect of one portfolio on another.

For definiteness, let us suppose the first portfolio is  $x$ , where as above

$$\ln x = \beta_0 \ln x_0 + \beta_1 \ln x_1 + \dots + \beta_n \ln x_n,$$

and the second portfolio is  $y$ , where

$$\ln y = \gamma_0 \ln y_0 + \gamma_1 \ln y_1 + \dots + \gamma_n \ln y_n.$$

We take  $\beta_0 = \gamma_0 = 1$ , and view  $\ln x_0$  and  $\varepsilon = \ln y_0$  as residuals with mean 0. The latter residual is not assumed to be a factor in our multivariate model. Consider the following typical “what-if”

question: Let  $A$  and  $B$  be given positive constants. If we know  $x \geq A$  at time  $T$ , what is the probability that  $y \geq B$  at time  $T$ ? We give two approaches to this problem, the first probably quicker, but possibly not as accurate, using a regression to avoid at least some Monte Carlo calculations.

### 5.1 “What-if”: An approach involving part regression, part Monte Carlo

We have  $\ln y \geq \ln B$  iff  $\ln y - \varepsilon \geq \ln B - \varepsilon$ . All correlations  $\rho_{ij}$  between  $\ln x_i$  and  $\ln x_j$  are assumed known. We may also assume that we have historical values of volatilities  $\sigma_i = \sqrt{\text{Var}(\ln x_i)}$ . (Alternatively, we could estimate such values as expected values of implied volatilities, but it would not be difficult to maintain an inventory of historical values, and more in the spirit of this part of the calculation to do so.) Thus we can estimate the historical covariances between  $\ln x$  and  $\ln y - \varepsilon$ :

$$\text{Cov}(\ln x, \ln y - \varepsilon) \approx \sum_{i,j} \beta_i \sigma_i \rho_{ij} \gamma_j \sigma_j T,$$

as well as  $\sigma_{\ln x} = \sqrt{\text{Var}(\ln x)}$ ,  $\sigma_{\ln y - \varepsilon} = \sqrt{\text{Var}(\ln y - \varepsilon)}$  and the correlation

$$\rho(\varepsilon) = \rho_{\ln x, \ln y - \varepsilon} = \frac{\text{Cov}(\ln x, \ln y - \varepsilon)}{\sigma_{\ln x} \sigma_{\ln y - \varepsilon}}.$$

This gives a standard regression for the variable  $\ln y - \varepsilon$  expressed in standard deviations from its mean, in terms of a similarly standardized expression for  $\ln x$ . Note that  $\varepsilon$  has mean 0 by construction. Put  $d_2(s^*, x, \sigma) = \frac{\ln(s^*/x) - \sigma^2 T/2}{\sigma \sqrt{T}}$ . Thus  $-d_2(s_x^*, x, \sigma_{\ln x})$  measures standardized  $\ln x$  using historical volatility, and  $-d_2(x) = -d_2(s_x^*, x, \hat{\sigma}_1(x))$  measures “standardized” (warped)  $\ln x$  using the cdf-implied volatility curve  $\hat{\sigma}_1(x)$ , as discussed in the previous section. Here  $s_x^*$  denotes our best estimate for the value of  $x$  at time  $T$ .

Let  $\sigma_1(y, \varepsilon)$  denote the volatility curve associated with  $\ln y - \varepsilon$ , which may be estimated as in the previous section (or computed from estimates of  $\sigma_1(y)$  and the standard deviation of the residual, if we are willing to view the residual as uncorrelated with  $\ln y - \varepsilon$ , as is guaranteed in unconstrained regression). Put  $d_2(y, \varepsilon) = d_2(s_y^*, y e^{-\varepsilon}, \hat{\sigma}_1(y, \varepsilon))$ , so that  $-d_2(y, \varepsilon)$  is a “standardized” measure of  $\ln y - \varepsilon$ . Then the standard regression appropriate to our model is

$$-d_2(y, \varepsilon) = \rho(\varepsilon)(-d_2(x)).$$

There is a residual associated with this regression, which we have not written down. It is presumably normal, and its variance may be computed. For notational reasons we will just imagine it has been incorporated into the original  $\varepsilon$ . As is apparent from the form of the expressions in the display, an alternative to the above regression is to do it with the warped correlation coefficients suggested in the previous section. If, in addition, it was appropriate to view the original portfolios as linear combinations of warped variables (our standard normal marginals), the regression above could be done without any recourse to Monte Carlo calculations. Similar remarks would apply if we used constant historical volatility functions throughout, though presumably the latter procedure would lose accuracy.

In any case, we can now answer our “what-if” question as a simple expectation in the univariate normal distribution of the (adjusted) residual  $\varepsilon$ . Abbreviate  $d_2(s_x^*, A, \hat{\sigma}_1(A))$  to  $d_2(A)$  and

$d_2(s_y^*, Be^{-\varepsilon}, \hat{\sigma}_1(B, \varepsilon))$  to  $d_2(B, \varepsilon)$ . Assume  $\rho(\varepsilon) > 0$  (the natural case of a positive correlation). Then we have

$$\begin{aligned}\Pr\{y \geq B \mid x \geq A\} &= \mathbb{E}(\Pr(-d_2(y, \varepsilon) \geq -d_2(B, \varepsilon) \mid -d_2(x) \geq d_2(A))) \\ &= \mathbb{E}(\Pr(-d_2(x) \geq \rho(\varepsilon)^{-1}(-d_2(B, \varepsilon)) \mid -d_2(x) \geq -d_2(A))) \\ &= \mathbb{E}(\min\{1, N(\rho(\varepsilon)^{-1}(d_2(B, \varepsilon))/N(d_2(A)))\}).\end{aligned}$$

The first equation follows just because  $-d_2(y, \varepsilon)$  is monotonically increasing as a function of  $y$ ; that is, the condition that  $y \geq B$  is completely equivalent to the condition  $-d_2(y, \varepsilon) \geq -d_2(B, \varepsilon)$ . Similar remarks hold for the condition  $x \geq A$ , while the expression  $\Pr\{y \geq B \mid x \geq A\}$  just means the probability that the condition  $y \geq B$  holds when it is known that  $x \geq A$ . The second equation is then derived with the displayed expression above for  $-d_2(y, \varepsilon)$ . (If  $\rho(\varepsilon)$  is negative, the inequality involving its inverse reverses.) This inner expectation is then calculated in the normal distribution. For values of  $\varepsilon$  for which  $-d_2(A)$  is as large as  $\rho(\varepsilon)^{-1}(-d_2(B, \varepsilon))$ , the expectation is a certainty, and yields the value 1. When  $-d_2(A)$  is smaller than  $\rho(\varepsilon)^{-1}(-d_2(B, \varepsilon))$ , its cumulative normal distribution value  $N(-d_2(A))$  is smaller than  $N(\rho(\varepsilon)^{-1}(-d_2(B, \varepsilon)))$ , and the probability  $1 - N(-d_2(A)) = N(d_2(A))$  that the standard normal variable  $z = -d_2(x)$  is at least  $-d_2(A)$  is smaller than the corresponding probability  $1 - N(\rho(\varepsilon)^{-1}(-d_2(B, \varepsilon))) = N(\rho(\varepsilon)^{-1}(d_2(B, \varepsilon)))$  that  $z$  be at least  $\rho(\varepsilon)^{-1}(-d_2(B, \varepsilon))$ . The ratio  $N(\rho(\varepsilon)^{-1}(d_2(B, \varepsilon))/N(d_2(A)))$ , which is the desired inner expectation, is thus smaller than 1, as is appropriate for a probability, conditional or not. If  $\rho(\varepsilon)$  is negative, similar reasoning leads instead to the expression  $\mathbb{E}(\max\{0, (N(d_2(A)) - N(\rho(\varepsilon)^{-1}(d_2(B, \varepsilon))))/N(d_2(A))\})$  for the desired conditional probability. Although the final answer in either case is an expectation (over  $\varepsilon$ ), it is essentially an integral that could be computed quickly with power series. (A very simple and accurate power-series expansion of  $N(z)$  is given on p. 252 of the book by Hull cited above.) Using that, one could determine by iterative methods what value of  $\varepsilon$  makes, say, the ratio  $N(\rho(\varepsilon)^{-1}(d_2(B, \varepsilon))/N(d_2(A)))$  equal to 1, and then integrate the ratio against the standard normal pdf from  $-\infty$  to the determined value of  $\varepsilon$ , in the  $\rho(\varepsilon) > 0$  case. Similar remarks apply if  $\rho(\varepsilon) < 0$ . (Note that, if  $\rho(\varepsilon) = 0$ , the variables  $\ln x$  and  $\ln y$  are uncorrelated, and the conditional probability  $\Pr\{y \geq B \mid x \geq A\}$  is the same as the unconditional probability  $\Pr\{y \geq B\}$ .)

All of the latter calculations can be done very fast. Of course, we have already used some Monte Carlo calculations to get this far, unless we are in the simplified context of constant volatility functions.

## 5.2 “What-if”: The full Monte

It is easy to say how we would compute an answer to the same “what-if” question, using our full joint probability distribution. We simply write

$$\Pr\{y \geq B \mid x \geq A\} = \mathbb{E}(\Pr(-d_2(y, \varepsilon) \geq -d_2(B, \varepsilon) \mid -d_2(x) \geq -d_2(A)))$$

and interpret  $\ln x$  in  $-d_2(x)$ , and  $\ln y - \varepsilon$  in  $-d_2(y, \varepsilon)$  in terms of their expansions in  $\ln x_0, \ln x_1, \dots, \ln x_n$ . To compute, say the inner expectation by a Monte Carlo calculation, we would generate a large number of random samples of multivariate standard normal vectors  $z$  with covariance matrix  $C$ . We then take the average, over the samples  $z$  which happen to satisfy  $z \geq -d_2(A)$ , of the function which is 1 when  $-d_2(y, \varepsilon) \geq -d_2(B, \varepsilon)$  and 0 otherwise. We have not experimented to see whether this method yields better answers than the regression procedure above. Nevertheless, it illustrates how we could approach more sophisticated “what-if” questions

that could not be easily treated by regressions. For example, suppose we believe that factor  $w$  will remain in a range  $C \leq w \leq D$ , and ask the same question about  $y$ , subject to the same condition on  $x$ . This is hard to formulate in terms of regression, and is simply not possible in terms of single-factor regression. However, it is easy to answer with the full distribution:

$$\Pr\{y \geq B \mid x \geq A, C \leq w \leq D\} = \\ \mathbb{E}(\Pr(-d_2(y, \varepsilon) \geq -d_2(B, \varepsilon) \mid -d_2(x) \geq d_2(A), -d_2(C) \leq -d_2(w) \leq -d_2(D))).$$

Finally, we may not want to work in the log domain, which, if we started with a fixed portfolio  $x = h_1 x_1 + h_2 x_2 + \dots + h_n x_n$  would force us into an approximation, as noted. But, working with the full distribution, we can phrase a condition  $x \geq A$  as  $h_1 x_1(y_1) + \dots + h_n x_n(y_n) \geq A$ , in the language of the first section where the vector of  $y$ 's plays the role of our vector  $z$  here. Monte Carlo calculations can now proceed as before, using log domain expressions or not for the other conditions.

## 6 “You gotta believe” questions

In the previous section we were focusing on an investor thinking about the value of his or her portfolio  $y$  in response to the change in a factor  $x$ . Conversely, an investor might want to know what the investment world looks like if a given stock or index  $y$  goes to a certain level  $B$  at time  $T$ . What is the expected value  $A$  at time  $T$  of another portfolio  $x$ , or simply of one of the factors  $x_i$ ? Our main plan is, upon input by the user that  $y$  is going to level  $B$ , to list several assets  $x_i$  or factors/indices  $x$  most highly correlated with  $y$  and their expected values with  $y$  at  $B$ .

It would also be possible to display a confidence interval for each selected asset or factor, and have other information about its new projected probability distribution readily available. We could also offer comparisons with the old projected probability distribution of  $x$ , where no assumptions on  $y$  is made. Finally, in some cases, where it was possible to explain much of the variance of  $y$  with just a few  $x_i$  (appearing in the regression of  $y$ ), we could list percentage increases/decreases of a portfolio of these  $x_i$  required to make  $B$  the expected value of  $y$ , based solely on its dependence on this portfolio. (For example, the coefficients in the portfolio could come from the regression of  $y$  with respect to all the  $x_i$ , or some new regression might be done, perhaps allowing user-defined constraints). It should be mentioned that medians or modes are alternatives to expected values (means) here and above; in any case users will need to be educated about the fact that the median and mode differ systematically from the mean in near lognormal distributions.

The main problem might be viewed as understanding the probability distribution of  $x$ , given that  $y \geq B$  at a given time  $T$ , with  $x$  and  $y$  as in the previous section. This can be approached by the methods of the previous sections, by reversing the roles of the variables.

There is, however, a simpler question that can be treated in an especially quick way. Consider the problem of determining the mean of  $x$  conditioned on the equality  $y = B$  at time  $T$ . The idea is to use simple regression methods, but interpret answers as measured in terms of our variable volatilities. In our previous notation, we have a regression

$$-d_2(x) = \rho \cdot (-d_2(y, \varepsilon)) + v$$

where  $\rho$  (which we called  $\rho(\varepsilon)$  in the previous section) is the historically determined correlation between  $\ln x$  and the random variable  $\ln y - \varepsilon$ . Note that the roles of dependent and independent

variable are reversed. There is also a residual  $v$ , which has mean 0 here, and plays no role (gets averaged away). Thus, the desired conditional expected value  $A$  of  $x$  is obtained from

$$-d_2(A) = E(-d_2(x) \mid y = B) = \rho \cdot E(-d_2(s_y^*, B e^{-\varepsilon}, \hat{\sigma}_1(y, \varepsilon))) = \rho \cdot (-d_2(s_y^*, B, \hat{\sigma}_1(y, \varepsilon))).$$

Recall that  $\hat{\sigma}_1(y, \varepsilon)$  is an estimate, obtained by Monte Carlo methods, of the implied volatility  $\sigma_1$  associated to the random variable  $\ln y - \varepsilon$ . For faster but less accurate calculations it can be estimated historically as  $\sum_{i,j} \beta_i \sigma_i \rho_{ij} \beta_j \sigma_j$  with each of the  $\sigma$ 's,  $\beta$ 's, and  $\rho$ 's here given historically. (See the previous section for notation.) Similarly, for fast calculations,  $-d_2(x)$  could use historical volatility, though we expect it to be given more accurately, or rather, more accurately according to the market view, as  $-d_2(x) = -d_2(s_x^*, x, \hat{\sigma}_1(x))$ , using the implied volatility function estimate  $\hat{\sigma}_1(x)$ . If  $x = x_i$  is a single asset or index in our model, then  $\hat{\sigma}_1(x) = \hat{\sigma}_1(x_i)$  does not require a Monte Carlo estimate, but is presumably already available.

To summarize, the conditional expected values required to answer “you gotta believe” questions are easily obtained by regression methods. The accuracy of such answers is enhanced, or at least shaped more to reflect market input, when all logvariables are measured in “standard deviations,” interpreted as our variable volatilities.

## 7 Portfolios containing option securities

We conclude this document by briefly pointing out that our methods, when using full Monte Carlo calculations, easily apply to portfolios containing option securities. The well-known idea is to think of an option as a kind of nonlinear portfolio—a quadratic one, to be more precise. Thus, an option on a single underlying security with underlying price  $x_1$  has a price approximately  $x = c + \Delta(x_1 - s_1) + (1/2)\Gamma(x_1 - s_1)^2$  for  $x_1$  near  $s_1$ , where the option was evaluated to a known value  $c$ . Here  $\Delta$  and  $\Gamma$  are well-known parameters in the options markets, giving the first and second derivatives of the option price at  $s_1$  with respect to the underlying security price  $x_1$ . Perhaps the most characteristic feature of options is that they have nonzero  $\Gamma$ —their proportion of increase or decrease with respect to the underlying security price changes as the security price changes. Explicit formulas in terms of other standard parameters are available, say, in the Black-Scholes theory for both  $\Delta$  and  $\Gamma$  (see the Hull book cited above). Such formulas could be obtained by differentiation directly in other theories or when using empirically-fitted curves. In any case, once we have such an explicit approximation to  $x$ , its probability distribution is easily given by the Monte Carlo methods of Subsection 3.1 above. The same method applies as well to portfolios containing several options and other securities.